Elucidating the Mechanism of mRNA Export Regulation by the Nuclear Pore Complex

Thesis by

Sarah Wang Cai

In Partial Fulfillment of the Requirements for the degree of

BACHELOR OF SCIENCE IN CHEMISTRY

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2018

Defended June 8, 2018

ACKNOWLEDGEMENTS

First, I would like to thank my thesis advisor, André Hoelz, for welcoming me into his group even when I was a freshman with no prior research experience. The past four years working in your lab have taught me so much and have been instrumental in my growth as a student and as a scientist. In your lab, I developed a passion for protein structure and function and for academic research at the highest level. I am deeply grateful for your endless support and confidence in me. Working with you has been the highlight of my college career.

To Dan Lin, I cannot thank you enough for your mentorship throughout the years. You have been my greatest teacher, and your extraordinary dedication to science and to mentorship has been a constant source of inspiration as I navigate my own scientific path. Thank you for your patience in teaching me structural biology from the ground up, for answering all my silly questions, and for giving me my own project and encouraging my independence from an early stage. That has really made all the difference. Thank you again for being such a phenomenal mentor and friend; I am honored to have been your student.

To Ana Correia, thank you for being an amazing collaborator, mentor, and friend. Thank you for taking me under your wing and teaching me how to work (and struggle) with RNA. To Chris Bley, although we have worked together for the shortest amount of time, thank you for all your help and wisdom. I have enjoyed tackling the challenge of the "Franken-ÄKTA" with you. To Ferdinand Huber, although you were never officially one of my mentors, you

have been such a great friend and teacher. Thank you for always being there for all my extra questions and for our many interesting conversations.

Thank you to the rest of the Hoelz lab for creating such a fun working environment. I had a blast working with you all, and thank you for welcoming me.

Finally, I would like to thank Mitchio Okumura for coordinating the senior thesis program and Bil Clemons for supporting me as an academic advisor.

ABSTRACT

The nuclear pore complex (NPC) is one of the largest and most complex protein assemblies in eukaryotic cells, and it facilitates and regulates the bidirectional transfer of molecules between the nucleus and cytoplasm. One of the essential functions of the NPC is to directly regulate the export of mature mRNAs, but the mechanism is not well understood, especially in humans. Export of mRNA is completed at the cytoplasmic side of the NPC, where ATPase activity of the DEAD-box helicase DDX19 is specifically activated by the NPC components Gle1, Nup42, and Nup214. The mRNA export factor Gle1 is an essential nucleoporin, and Gle1 dysfunction has been linked to human diseases. We show that the Gle1•Nup42 interaction is highly conserved with X-ray crystal structures and that the thermostability of Gle1 is highly dependent on Nup42. We also find that disease-linked mutants of Gle1 show strongly altered thermostability. Analysis of DDX19 steady-state ATPase activity reveals a novel mode of activation by Gle1 in humans. Structural studies of DDX19 complement the biochemical characterization of DDX19, and we have proposed a working model for the DDX19 catalytic cycle. An outstanding question remains in this model, as the details of DDX19-mediated removal of the export factor NXF1•NXT1 from mRNA are unknown. We describe a method for large-scale purification of retroviral constitutive transport element (CTE) RNA, which hijacks NXF1•NXT1 to be exported through the NPC, for X-ray crystallographic studies of a NXF1•NXT1•CTE complex. Together, these results provide a detailed structural and functional description of mRNA export regulation by the NPC and a framework for understanding the molecular basis of human disease linked to Gle1 and viral proliferation.

PUBLISHED CONTENT AND CONTRIBUTIONS

Daniel H. Lin, Ana R. Correia*, **Sarah W. Cai***, Ferdinand M. Huber, Claudia A. Jette, André Hoelz (2018). Structural and functional analysis of mRNA export regulation by the nuclear pore complex, ***Nature Communications***, accepted.

Contributions: cloning, protein expression and purification, protein crystallization, structure determination of scGle1^{CTD}•Nup42^{GBM}, hsGle1^{CTD}•Nup42^{GBM}, yeast analyses, protein-protein interaction experiments, thermostability experiments, ATPase activity experiments

*denotes co-second authors

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

INTRODUCTION

A hallmark of the evolution separating eukaryotes from bacteria is the subdivision of cells into specialized, membrane-bound compartments. One such membrane bilayer is found in the nuclear envelope, separating genetic information contained in the nucleus from the greater cell cytoplasm. The nuclear pore complex (NPC) is nature's largest protein channel, and it is the sole gateway for the bidirectional transport of macromolecules between the nucleus and cytoplasm of eukaryotic cells. Although the NPC is massive, with a diameter of ~1000 Å and a total mass of ~120 MDa, the eightfold-symmetric complex is only composed of approximately 30 unique nuclear pore proteins (nucleoporins) which arrange into a symmetric, doughnut-shaped core and asymmetric nuclear basket and cytoplasmic filament subcomplexes (Hoelz et al., 2011) (Figure 1). Whereas the composite structure of the symmetric core of the NPC has recently been elucidated, the architecture of the asymmetric nuclear and cytoplasmic faces is yet to be determined (Kosinski et al., 2016; Lin et al., 2016). Here, we focus on studying the structure and function of the mRNA export machinery of the NPC, located at the cytoplasmic face (circled in Figure 1).

The central dogma of molecular biology poses that genetic information flows from DNA to protein in a two-step process. First, DNA is converted to RNA through the process of transcription, and then protein is synthesized from RNA in the process of translation. The precisely controlled handoff between transcription and translation is a fundamental biological process that is necessary for all cellular function. In eukaryotes, transcription and translation are spatially and temporally segregated in the nucleus and cytoplasm,

respectively. Whereas prokaryotes couple the two processes, the separation allows eukaryotes to have far greater control and regulation of gene expression, contributing to the domain's greater diversity of form and function (Hoelz et al., 2011).

Messenger RNA (mRNA), the template for protein synthesis, is transcribed from DNA in the nucleus. After post-transcriptional processing, such as splicing and addition of a 5' cap and poly(A)+ tail, the mature mRNA is packaged with additional proteins into an export-competent messenger ribonucleoprotein (mRNP) complex (Stewart, 2010). This mRNP interacts with and shuttles through the NPC diffusion barrier formed by phenylalanine glycine (FG) repeats in the central channel of the pore (Katahira et al., 1999). At the cytoplasmic face of the NPC, the mRNP encounters machinery that releases the mRNA into the cytoplasm for translation, but the molecular mechanism of this mRNP remodeling event remains to be elucidated (Lund and Guthrie, 2005).

This work describes efforts to characterize the cytoplasmic-side mRNA export machinery of the NPC using an interdisciplinary approach, combining structural studies and biochemical analyses. The first chapter explores the interaction between the mRNA export factor Gle1 and the cytoplasmic filament nucleoporin Nup42, which was previously uncharacterized. The molecular details of the Gle1•Nup42 interaction are set into the greater context of mRNA export regulation by the NPC in the second chapter. Stimulation of the mRNA export factor DDX19, the catalytic center of the export machinery, is reconstituted *in vitro*, providing new insights into the role of nucleoporins in the termination of mRNA export. Finally, the third chapter describes the crystallography-scale synthesis and purification of the viral constitutive transport element (CTE) RNA, which provides a

framework for the structural study of the specific interaction between the RNA-binding heterodimeric export factor NXF1•NXT1 and CTE-RNA that can advance knowledge of the precise mechanism for mRNP remodeling by DDX19. Together, these results provide a detailed structural and functional description of mRNA export regulation by the NPC and a framework for understanding the molecular basis of human disease linked to Gle1 and viral proliferation.

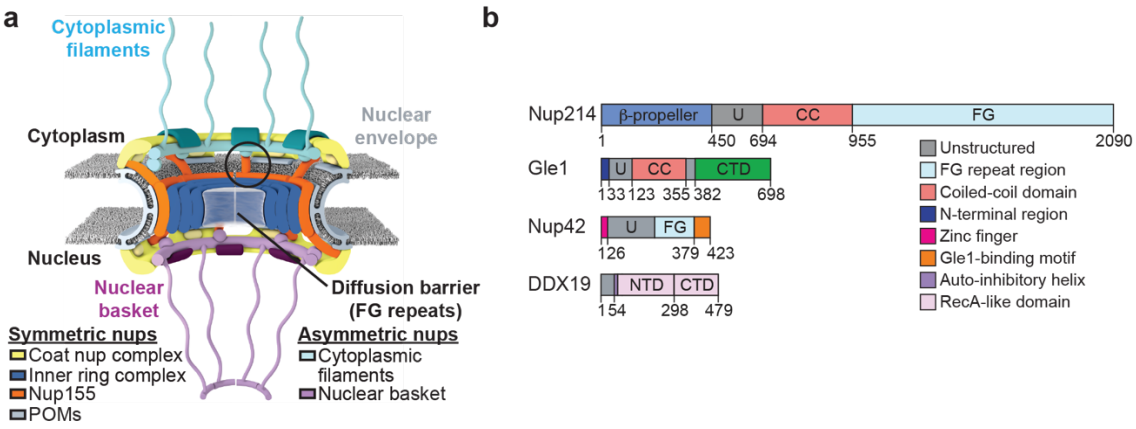


Figure 1. The mRNA export machinery of the nuclear pore complex. (a) Cartoon schematic of the human nuclear pore complex (NPC). The circle highlights the region of the NPC to which the proteins used in this study localize. (b) Domain schematics for nucleoporins used in this study. Protein names and boundaries correspond to the human proteins (Adapted with permission from Lin et al., 2018).

Chapter 1

STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF THE GLE1•NUP42 INTERACTION

This chapter contains results generated by colleagues to confer a coherent story and flow of logic. Experiments conducted by others will be explicitly declared in blue. This chapter was adapted from:

Daniel H. Lin, Ana R. Correia*, **Sarah W. Cai***, Ferdinand M. Huber, Claudia A. Jette, André Hoelz (2018). Structural and functional analysis of mRNA export regulation by the nuclear pore complex, ***Nature Communications***, accepted April 27, 2018.

Introduction

Gle1 is an essential mRNA export factor that is required for poly(A)+ mRNA export, in addition to its other functional roles in the cell. The majority of Gle1 studies have been conducted in budding yeast, *S. cerevisiae* (sc), where Gle1 interacts with multiple binding partners. One of these partners is Dbp5 (DDX19 in humans), which is a conserved essential DEAD-box family helicase (Kendirgi et al., 2003; Montpetit et al., 2011; Rollenhagen et al., 2004). The association of the Gle1 C-terminal domain ($Gle1^{CTD}$) and the Dbp5 C-terminal domain ($Dbp5^{CTD}$) is mediated by the small molecule inositol hexaphosphate (IP_6), and the structure of this complex has been solved (Montpetit et al., 2011). Gle1 is also known to interact with Nup42, a cytoplasmic filament nucleoporin, via their C-terminal regions, but otherwise, the role of the interaction between Gle1 and Nup42 is not well understood (Kendirgi et al., 2005). Previous studies in *S. cerevisiae* have shown that Gle1 is able to localize to the nuclear envelope in a *nup42*Δ knockout strain at 23°C. However, after heat shock to 42°C, Gle1 mislocalizes and the cells experience defects in mRNA export, suggesting that Nup42 strengthens the interaction between Gle1 and the NPC but is only necessary during heat shock (Rollenhagen et al., 2004). In humans (hs), Gle1 exists in two isoforms, as a result of two differing splice variants. Gle1A lacks 43 residues at the C-terminus that are required for Nup42 binding, and in the cell, this isoform remains cytoplasmic. The more abundant Gle1B, homologous to scGle1, is capable of binding Nup42 and localizes to the NPC (Kendirgi et al., 2005). Throughout this text, Gle1 will refer to the Gle1B isoform of hsGle1.

In humans, the Gle1•Nup42 interaction is particularly interesting in the context of human health and disease. A recent study has linked three specific mutations in Gle1 to amyotrophic lateral sclerosis (ALS), and other mutations identified in Gle1 have been linked to lethal congenital contracture syndrome (LCCS1) and lethal arthrogryposis with anterior horn cell disease (LAAHD) (Kaneb et al., 2015; Nousiainen et al., 2008). LCCS1 and LAAHD are autosomal recessive motor neuron disorders which result in fetal death before the 32nd gestational week, and LCCS1 is the more phenotypically severe of the two. In both diseases, patients have a characteristic three amino acid (PFQ) insertion at residue 144 in the coiled-coil region of Gle1, termed Fin_{Major}. LCCS1 patients are homozygous for the Fin_{Major} mutation, while LAAHD patients are compound heterozygous for the Fin_{Major} mutation and a single amino acid substitution in $Gle1^{CTD}$ (V617M and I684T). In a Fin_{Major} heterozygous patient diagnosed with LCCS1 based on phenotypic similarities, a third mutation (R569H) was identified, suggesting that the R569H has a stronger effect compared to the V617M and I684T substitutions (Nousiainen et al., 2008). ALS is a fatal neurodegenerative disease that has been recently characterized as an RNA metabolism disorder. The three mutations identified in Gle1 consist of a nonsense mutation that truncates the protein at residue 70, a splice-site mutation that replaces the C-terminal 44 residues of Gle1 with a novel 88-residue stretch, and a missense mutation that results in the substitution R697C. The latter two mutations both affect the Nup42 binding domain of Gle1, which is the interaction of interest presented here (Kaneb et al., 2015). Understanding the effects of these mutations on Gle1 in the context of the NPC and mRNA export will help increase our understanding of the molecular-level causes of motor neuron diseases.

In this chapter, we present a structural and functional analysis of the role of Nup42 as a binding partner for Gle1, and we examine the effect of disease mutations on Gle1 function.

Crystallization of yeast and human Gle1•Nup42

To study the molecular details of the Gle1•Nup42 interaction, we reconstituted the complex *in vitro* with purified, recombinant proteins and pursued an X-ray crystal structure to study the interaction at atomic resolution. First, we sought to determine the minimal span of Nup42 required to interact with $Gle1^{CTD}$, which refers to the previously solved domain containing residues 244-538 (Montpetit et al., 2011) and the analogous residues (382-698) in the human protein (domain boundaries are schematized in introduction Figure 1 and Figure S1). The rationale for this experiment lies in the hypothesis that residues of Nup42 at the $Gle1^{CTD}$-interaction interface will be structured, eliminating any disordered regions that would interfere with crystallization attempts. Using constructs of Nup42 fused to a glutathione S-transferase (GST) affinity tag and $Gle1^{CTD}$ fused to SUMO, we overexpressed yeast and human GST-Nup42 and SUMO-$Gle1^{CTD}$ in *E. coli*, pulling down SUMO-$Gle1^{CTD}$ from the bacterial lysate (Figure S1). The minimal $Gle1^{CTD}$-binding region on Nup42 was mapped to residues 397-430 in yeast and the analogous residues 379-423 in humans, termed the Gle1-binding motif (GBM) (Introduction Figure 1). Preliminary *in vivo* localization experiments in *S. cerevisiae* confirmed that $Nup42^{GBM}$-mCherry was sufficient for rescue of heat shock induced mislocalization of Gle1-GFP (Figure S2). The minimal $Nup42^{GBM}$ constructs were coexpressed and purified with $Gle1^{CTD}$. Crystals of sc$Gle1^{CTD}$•$Nup42^{GBM}$ were obtained in

the tetragonal space group $P4_32_12$. The structure was solved using molecular replacement and refined to 1.75 Å resolution with an R_{work}/R_{free} of 18.5/21.1.Crystals of hsGle1^{CTD}•Nup42^{GBM} were obtained in the monoclinic space group C2. The structure was solved using molecular replacement and refined to 2.8 Å resolution with an R_{work}/R_{free} of 24.5/27.4. Details of data collection and refinement statistics are given in Table 1.

Nup42 interacts with Gle1 via a conserved binding motif

The structures of human and yeast Gle1^{CTD}•Nup42^{GBM} show remarkable similarity, especially when considered in the context of a third structure of the thermophilic fungus *C. thermophilum* (ct) Gle1^{CTD}•Nup42^{GBM}. In all three species, Nup42^{GBM} folds into a compact helical secondary structure that buries a solvent-exposed hydrophobic surface on Gle^{CTD}, which also adopts the same fold in all three species (Figure 1). In yeast, scNup42^{GBM} residues F409, F414, P420, and P423 appear to be critical for folding of the Nup42^{GBM} hydrophobic core around scGle1^{CTD} residues W451 and Y488. These key residues are conserved across the three species (hsNup42^{GBM} residues F401/F406/P412/P415 recognize hsGle1^{CTD} residues W602/Y637; ctNup42^{GBM} residues F533/F539/P544/P547 recognize ctGle1^{CTD} residues W447/Y484) (Figure 1e). The structural similarities are consistent with a multi-species sequence alignment of Gle1^{CTD}, in which the residues involved in Nup42-binding are highly conserved (Figure S3). The persistence of this interaction through evolution supports an important functional role for the complex. To validate the structure, we generated variants of scNup42^{GBM} containing mutations in key residues from the interface: alanine substitutions F409A, F414A, L416A, P423A and charge-introducing substitutions L416R, F409D,

F414D, and a F409D/F414D double mutant. While single alanine substitutions in the hydrophobic core did not abolish binding to scGle1^{CTD}, consistent with the extensive interaction surface, the highly charged aspartate substitutions (F414D and F409D/F414D) were able to disrupt the interaction on a SEC column (Figure 2).

Gle1 thermostability is dependent on Nup42

During the expression and purification of Gle1^{CTD}•Nup42^{GBM}, we observed that *apo* Gle1^{CTD} had far lower expression yields than the complex and was unstable and prone to aggregation. In yeast, addition of IP_6 to the purification buffers reduced precipitation, but the yield was still far lower than that obtained with coexpressed scGle1^{CTD}•Nup42^{GBM}. With the human proteins, *apo* hsGle1^{CTD} expression yields were even more abysmal, and IP_6 addition did not improve stability during purification. Taking into context the temperature-sensitive role of Nup42 in Gle1 localization to the NPC and the hydrophobic nature of the interaction interface, we tested the effects of Nup42^{GBM} and IP_6 on Gle1 thermostability. Using two complementary methods, differential scanning fluorimetry (DSF) and a protein solubility assay, we measured the effects of Nup42^{GBM} and IP_6 on the melting temperature (T_m) of Gle1^{CTD}. In yeast, we observed a dramatic effect for each of IP_6 and scNup42^{GBM}, with shifts in T_m from 22°C for *apo* scGle1^{CTD} to 37°C and 46°C, respectively. The combined effect was even greater, shifting the T_m to 53°C. In humans, a similar effect was observed for hsNup42^{GBM}, which shifted the T_m from 37°C for *apo* hsGle1^{CTD} to 50°C. When copurified with hsNup42^{GBM}, expression yields increased 100-fold for the complex over *apo* hsGle1^{CTD}. IP_6 did not have a significant effect on human Gle1^{CTD} thermostability. Remarkably,

apo Gle1^{CTD} is unstable below the physiological growth temperatures of 30°C in yeast and 37°C in humans (Figure 3). These results suggest an important role for cofactors in Gle1 thermostability and suggest that the main role of Nup42^{GBM} is to ensure proper folding of Gle1, consistent with the heat shock-dependent mislocalization of Gle1 *in vivo*.

Gle1 disease mutants have altered thermostability

We next generated the hsGle1^{CTD} point mutants implicated in disease (LCCS1 mutation R569H, LAAHD mutations V617M and I684T, and ALS mutation R697C) to investigate the functional effect of the mutations on hsGle1^{CTD}•Nup42^{GBM} thermostability. Strikingly, all three LAAHD/LCCS1 mutants displayed a lower T_m relative to WT hsGle1^{CTD}•Nup42^{GBM}, with the R569H mutant showing the greatest defect in thermostability with a shift from 50°C to 40°C, correlating well with its more severe phenotype. Although the ALS point mutation did not appear to have an effect on thermostability, the LAAHD/LCCS1 mutant analysis suggests a possible role for Gle1-misfolding in disease (Figure 4).

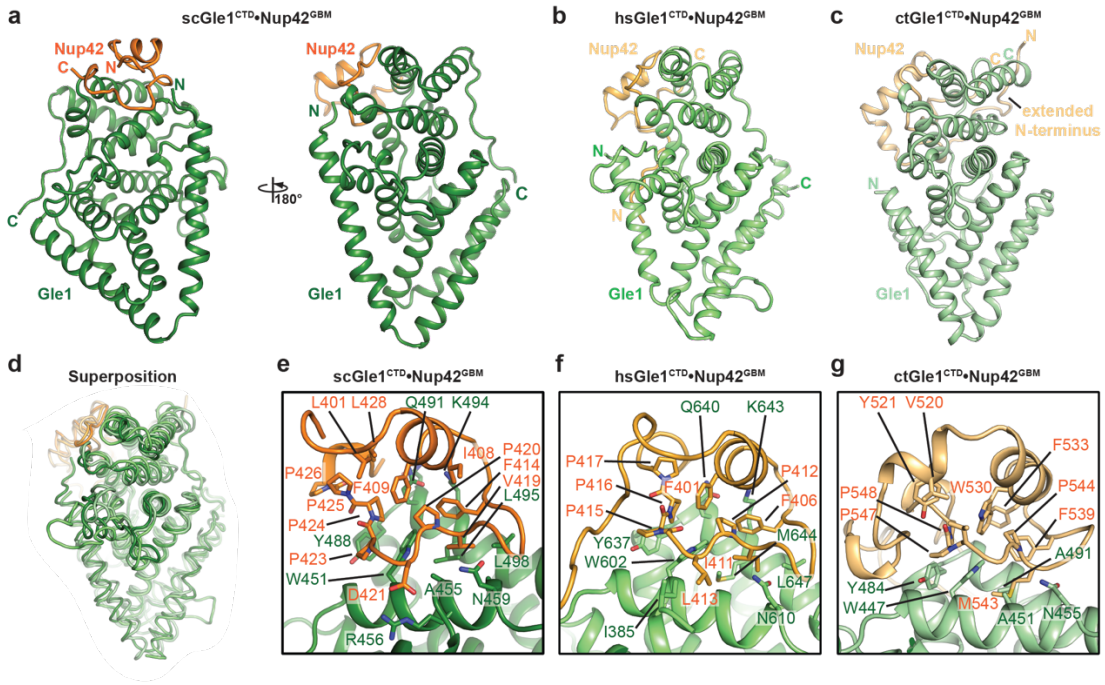


Figure 1. A conserved mechanism for Gle1•Nup42 complex formation. Experimental data presented in panels c and g of this figure have been conducted by A.R.C. (a-c) Crystal structures of (a) *S. cerevisiae,* (b) *H. sapiens*, or (c) *C. thermophilum* Gle1^{CTD}•Nup42^{GBM}. (d) Superposition of the structures of *S. cerevisiae, H. sapiens*, and *C. thermophilum* Gle1^{CTD}•Nup42^{GBM}, with same coloring as in (a-c). (e-g) Zoom views of (e) *S. cerevisiae,* (f) *H. sapiens*, or (g) *C. thermophilum* Gle1^{CTD}•Nup42^{GBM} interactions with residues mediating the interaction labeled.

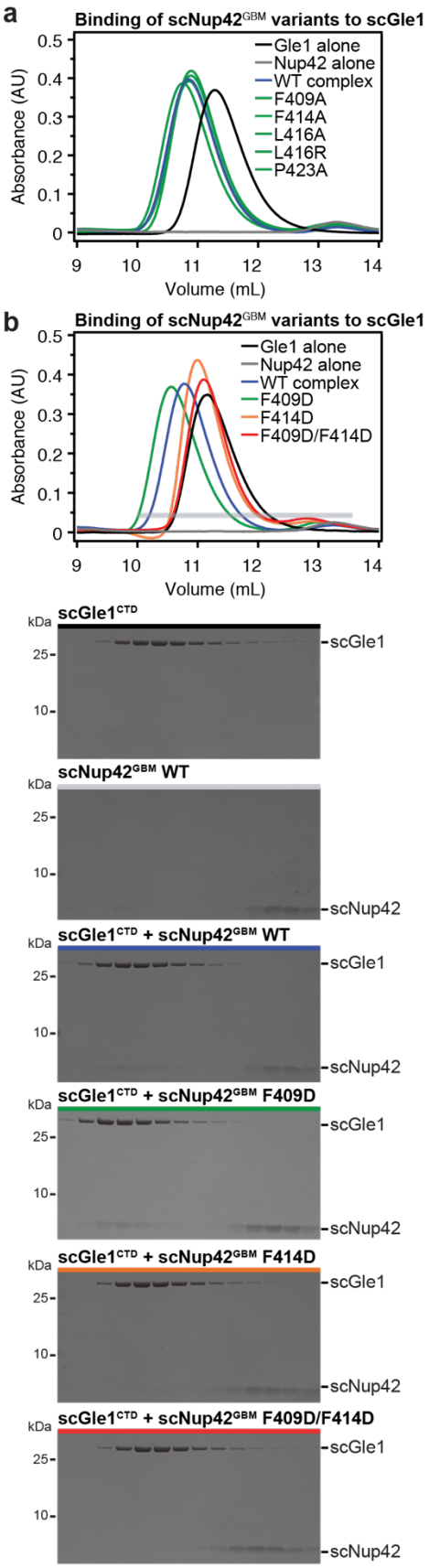


Figure 2. Mutational analysis of *S. cerevisiae* Gle1^{CTD}•Nup42^{GBM}. This experiment was conducted together with D.H.L.; S.W.C. generated the mutants for analysis. (a, b) SEC analysis of the effect of mutations in Nup42^{GBM} on Gle1^{CTD} binding. Purified Gle1^{CTD} was mixed with the indicated Nup42GBM mutants and loaded on a Superdex 75 10/300 GL column. SEC profiles of Gle1^{CTD} (black), Nup42^{GBM} (grey), and Gle1^{CTD} preincubated with Nup42^{GBM} (blue) are shown as controls. SEC profiles of Nup42^{GBM} mutants preincubated with Gle1^{CTD} are colored green for wild-type levels of complex formation, orange for reduced binding, or red for complete disruption. The gray bar indicates the fractions visualized with Coomassie-stained SDS-PAGE gels.

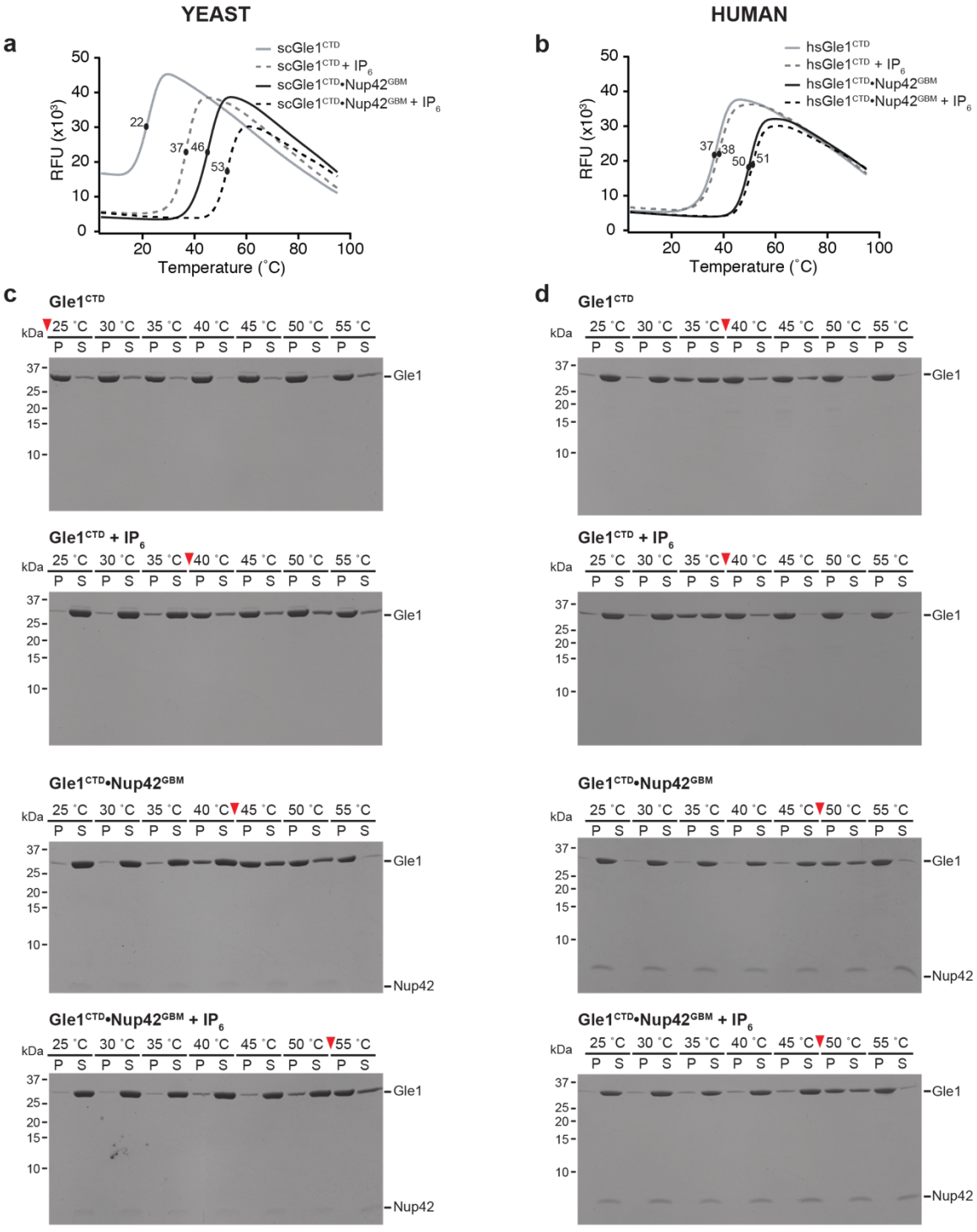

YEAST
HUMAN
a
b
scGle1CTD
scGle1CTD + IP6
scGle1CTD•Nup42GBM
scGle1CTD•Nup42GBM + IP6
hsGle1CTD
hsGle1CTD + IP6
hsGle1CTD•Nup42GBM
hsGle1CTD•Nup42GBM + IP6
RFU (x10³)
Temperature (˚C)
22
37
46
53
37
38
50
51
c
d
Gle1CTD
Gle1CTD + IP6
Gle1CTD•Nup42GBM
Gle1CTD•Nup42GBM + IP6
kDa
25 ˚C
30 ˚C
35 ˚C
40 ˚C
45 ˚C
50 ˚C
55 ˚C
P S
Gle1
Nup42

Figure 3. Gle1 thermostability analysis. (a, b) Differential scanning fluorimetry analysis of (a) *S. cerevisiae* or (b) *H. sapiens* $Gle1^{CTD}$ thermostability in the presence and absence of $Nup42^{GBM}$ and IP_6. Exposure of hydrophobic residues was monitored by an increase in relative fluorescence units (RFUs). Curves represent the average of three experiments. (c-d) (c) *S. cerevisiae* or (d) *H. sapiens* $Gle1^{CTD}$ and $Gle1^{CTD}$•$Nup42^{GBM}$ were incubated at the indicated temperatures in the absence or presence of IP_6 for 30 minutes prior to centrifugation. Pelleted (P) and soluble (S) fractions were analyzed by SDS-PAGE and visualized by Coomassie staining. Red arrows indicate the temperature at which more than 50% of total $Gle1^{CTD}$ pelleted.

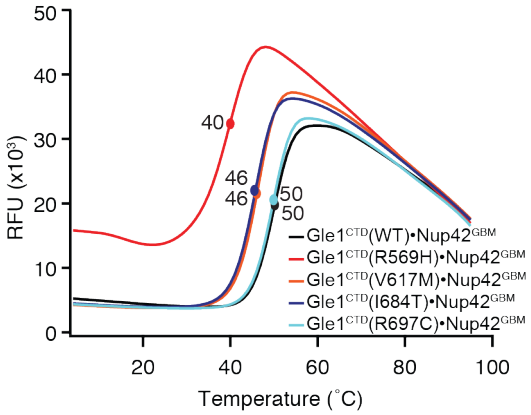


Figure 4. Thermostability of human disease variants of Gle1^{CTD}. Exposure of hydrophobic residues was monitored by an increase in relative fluorescence units (RFUs). Curves represent the average of three experiments.

Table 1.

Data collection and refinement statistics for *S. cerevisiae* and *H. sapiens* $Gle1^{CTD}$•$Nup42^{GBM}$structures

	$scGle1^{CTD}$•$Nup42^{GBM}$	$hsGle1^{CTD}$•$Nup42^{GBM}$
Data collection		
Space group	$P4_32_12$	C2
Cell dimensions		
a, *b*, *c* (Å)	64.5, 64.5, 361.7	163.7, 69.2, 93.0
α, β, γ (°)	90.0 90.0 90.0	90.0, 90.6, 90.0
Resolution (Å) [a,b]	50.0–1.75 (1.81–1.75)	50.0–2.8 (2.9–2.8)
R_{meas} [b]	12.4 (127.0)	10.7 (144.3)
$I / \sigma I$ [b]	16.4 (1.3)	10.1 (1.1)
Completeness (%) [b]	99.3 (93.5)	99.2 (97.8)
Redundancy [b]	21.8 (10.5)	3.8 (3.5)
Refinement		
Resolution (Å)	47.3–1.75	46.8–2.8
No. reflections	85,335	20,826[c]
R_{work} / R_{free}	18.5/21.1	24.5/27.4
No. atoms	7,314	5,835
Protein	6,674	5,794
Ligand/ion	8	12
Water	632	29
B-factors	39	59
Protein	39	59
Ligand/ion	43	63
Water	43	31
R.m.s. deviations		
Bond lengths (Å)	0.007	0.003
Bond angles (°)	1.1	0.6

[a]Diffraction data were obtained from a single crystal

[b]Values in parentheses are for highest-resolution shell.

[c]Refinement was performed with ellipsoidally truncated data

Chapter 2

BIOCHEMISTRY OF DDX19 ACTIVATION BY THE NPC

This chapter contains results generated by colleagues to confer a coherent story and flow of logic. Experiments conducted by others will be explicitly declared in blue. This chapter was adapted from:

Daniel H. Lin, Ana R. Correia*, **Sarah W. Cai***, Ferdinand M. Huber, Claudia A. Jette, André Hoelz (2018). Structural and functional analysis of mRNA export regulation by the nuclear pore complex, ***Nature Communications***, accepted April 27, 2018.

Introduction

The DEAD-box family helicase DDX19 (Dbp5 in fungi) is an ATPase that binds mRNA and is responsible for the terminal catalytic step in mRNA export by the NPC. DEAD-box proteins are characterized by their RNA-dependent ATPase activity and the presence of two RecA-like domains ($DDX19^{NTD}$ and $DDX19^{CTD}$) (Weirich et al., 2006; Dossani et al., 2009; Montpetit et al., 2011). DDX19/Dbp5 exhibits very weak ATPase activity alone, and it requires RNA and stimulation by nucleoporins for function (Alcázar-Román et al., 2006; Montpetit et al., 2011; Weirich et al., 2006). The major binding partner and activator for DDX19/Dbp5 is Gle1, and the two proteins interact via their C-terminal domains (Alcázar-Román et al., 2006; Montpetit et al., 2011; Weirich et al., 2006; Noble et al., 2011). From our structures of $Gle1^{CTD}$•$Nup42^{GBM}$, we know that the $Nup42^{GBM}$ interaction surface is spatially distinct from the surface on $Gle1^{CTD}$ recognized by DDX19/Dbp5. In yeast, IP_6 is required to form a complex between $Gle1^{CTD}$ and $Dbp5^{CTD}$, and it is also required for Gle1-mediated stimulation of Dbp5 ATPase activity (Alcázar-Román et al., 2006; Montpetit et al., 2011; Weirich et al., 2006; Noble et al., 2011).

Although reconstitution of fully activated human DDX19 has not yet been achieved, *in vitro* studies of the yeast proteins have revealed a combined stimulatory effect of Gle1, IP_6, and RNA on the steady-state ATPase activity of Dbp5. The precise mechanism by which these factors activate the enzyme is unclear, especially in humans (Noble et al., 2011; Montpetit et al., 2011). Furthermore, the roles of the cytoplasmic filament nucleoporins Nup42 and Nup214 (Nup159 in yeast) in the DDX19 catalytic cycle are not well understood. As discussed in the previous chapter, Nup42 functions in stabilizing the $Gle1^{CTD}$ fold, but

whether it directly affects DDX19 activity is unknown. The N-terminal domain of Nup214 is required for localization of DDX19 to the NPC, but its role in DDX19 catalytic activity is not well understood (Schmitt et al., 1999; von Moeller et al., 2009; Napetschnig et al., 2009). Nup214/Nup159 have been shown to inhibit RNA-mediated stimulation of DDX19/Dbp5, but the addition of Gle1 in both species rescues this inhibition, so the precise functional role of Nup214 also remains to be determined (von Moeller et al., 2009; Napetschnig et al., 2009).

Here, we report reconstitution of a fully stimulated DDX19 system *in vitro*, revealing an IP_6-independent mechanism that appears to be conserved in metazoans. Taking the biochemical data presented here together with concurrent X-ray crystallographic studies, we discovered the precise mechanism by which Gle1 activates DDX19 in humans and a novel stimulatory role for Nup214 in the context of Gle1-mediated DDX19 activation.

Reconstitution of DDX19 activity *in vitro*

To assay the stimulatory effect of NPC components on DDX19 activation in humans, we modified a previously described colorimetric NADH-coupled ATPase assay designed for Dbp5 (Montpetit et al., 2012). Human DDX19 exhibited much weaker activity than Dbp5, so the human protein concentrations were increased by 5-fold to generate enough signal to fit the NADH colorimetric decay. Compared to the yeast proteins, we observed a similar role for RNA and $Gle1^{CTD}$ on DDX19 stimulation, where each stimulates the enzyme individually and together exert a greater combined effect (Figures 1 and S4). The addition of $Nup42^{GBM}$ only conferred a modest additional stimulatory effect compared to *apo* $Gle1^{CTD}$ in both yeast and humans, which appeared to be related to the thermostability of $Gle1^{CTD}$, as the assays were conducted at temperatures greater than the T_m of *apo* $Gle1^{CTD}$. The effect of $Nup42^{GBM}$

was more pronounced when the assay was conducted at 37°C for both yeast and human proteins, further suggesting that Nup42^{CTD} does not directly stimulate DDX19 (Figures 1 and S4). Rather, it acts allosterically through stabilization of the Gle1 fold.

Gle1-mediated DDX19 activation is independent of IP_6 in humans

In yeast, IP_6 is required to both form the Gle1•Dbp5 complex and to observe Gle1-dependent stimulation of Dbp5 (Alcázar-Román et al., 2006; Montpetit et al., 2011; Weirich et al., 2006; Folkmann et al., 2011). Strikingly, we found that while RNA and Gle1 activated DDX19 similarly in both humans and yeast, IP_6 did not appear to be necessary for Gle1-mediated stimulation of DDX19 in the human system. IP_6 is not required for complex formation between hsGle1 and DDX19 (Figure S5), and Gle1^{CTD} stimulated DDX19 to the same level with and without IP_6 present (Figure 1). Furthermore, the IP_6 effect in yeast is concentration dependent, saturating around 1 μM (2-fold excess over Dbp5), whereas in humans, there is no observable stimulation of DDX19 up to 100 μM IP_6, the upper bound for total cellular IP_6 concentration (Figure 2a-b) (Shears, 2001).

Further insights into the molecular basis of this differential dependence on IP_6 come from examining more closely the structures of Gle1^{CTD}. IP_6 has six phosphate groups, harboring a strong total negative charge. Studying the surface electrostatic potential of the Dbp5/IP_6 interaction interface on Gle1^{CTD}, we find that in both *S. cerevisiae* and *C. thermophilum*, there exists a highly charged basic pocket that readily accommodates the IP_6 molecule. However, the analogous surface on hsGle1^{CTD} does not harbor the same electrostatic potential (Figure 2c-g). A multispecies sequence alignment suggests that IP_6

dependence is unique to fungi, as the conservation of charged residues implicated in IP_6-binding only extend through fungal species, not through metazoan species (Figure S5). Whether IP_6 dependence was lost in metazoans or uniquely gained in fungi remains to be studied.

Gle1 activates DDX19 via relief of auto-inhibition

We then probed the mechanism of how Gle1 stimulates DDX19, supplementing structural and sequence data with biochemistry. X-ray crystal structures of hsGle1^{CTD}•Nup42^{GBM}•DDX19$^{\Delta N53}$ in complex with ADP and with AMP-PMP•Mg^{2+}, a non-hydrolysable ATP analog, were obtained at 3.6 Å and 3.4 Å resolution, respectively. The construct of DDX19$^{\Delta N53}$ used has 53 residues of the DDX19 flexible N-terminal extension removed to facilitate crystallization (Figure S7). From these structures, we noted a major difference between DDX19 and the previously solved structure of scGle1^{CTD}•IP_6•Dbp5$^{\Delta N90}$ (Montpetit et al., 2011). In humans, DDX19 contains an N-terminal auto-inhibitory helix that binds tightly between the two RecA domains (Collins et al., 2009) in the same way regardless of which nucleotide is present. This auto-inhibitory helix is only present in metazoan species, as determined by sequence conservation analysis. In the crystal structure of scGle1^{CTD}•IP_6•Dbp5$^{\Delta N90}$, the lack of an auto-inhibitory helix appears to allow for a greater rotation of the Dbp5 N-terminal domain toward Gle1^{CTD} such that Dbp5^{NTD} also makes contact with Gle1^{CTD}, in addition to the primary interaction interface between Gle1^{CTD} and the Dbp5 C-terminal domain. In the human structure, DDX19^{NTD} is rotated away from Gle1 (Figure S7).

To describe the specific mechanism by which Gle1^{CTD} stimulates DDX19, we compared the structure of *apo* DDX19$^{\Delta N53}$, obtained at 2.2 Å resolution, to Gle1^{CTD}•Nup42^{GBM}-bound DDX19. We noted several conformational changes induced by Gle1^{CTD}-binding: a set of loop movements and rearrangements in DDX19^{CTD} results in lost interactions between DDX19^{CTD} and the auto-inhibitory helix, ultimately causing a partial separation of DDX19^{NTD} and DDX19^{CTD} in a manner similar to the full separation seen in the yeast structure (Figure S8). Due to likely thermodynamic favorability of the auto-inhibited DDX19, our structure represents an early, partially open state created by Gle1-binding. We hypothesize the fully open state, where DDX19^{NTD} contacts Gle1^{CTD} similar to in the yeast structure, also occurs in humans and is required for DDX19 activation. In a previous study of the yeast proteins, a triple amino acid substitution (V513D/A516D/I520D) at the interface between Dbp5^{NTD} and Gle1^{CTD} abolished Gle1-mediated stimulation of Dbp5 (Montpetit et al., 2011). We made the homologous triple mutation (G666D/I669D/Q673D, DDD mutant) in human DDX19 and observed the same detrimental effect on Gle1-mediated DDX19 stimulation (Figure S4g).

Thus, the structure-based hypothesis for how Gle1^{CTD} stimulates DDX19 activity poses that Gle1^{CTD} triggers opening of the DDX19 RecA domains such that the nucleotide (ADP or ATP) can be exchanged and the system recycled. To test this hypothesis, we made several variants of DDX19 for biochemical analysis: DDX19$^{\Delta N53}$ is the crystallized construct: DDX19$^{S60D/K64D}$ is mutated in two residues that are critical for interaction between the auto-inhibitory helix and DDX19^{NTD}, DDX19$^{\Delta N67}$ has the auto-inhibitory helix fully truncated, and DDX19$^{\Delta N91}$ further removes mobile residues, including the auto-inhibitory

helix (Figure 3a). These mutants all exhibited higher basal ATPase activity than DDX19^{WT}, suggesting that removal of the auto-inhibitory helix is a rate limiting step in the reaction. Furthermore, the fully stimulated levels of the hyperactive mutants approached yeast Dbp5 activity levels, suggesting the presence of the auto-inhibitory helix was a reason for the low activity observed with the human proteins. Addition of RNA suppressed the hyperactive mutants back down to WT levels, so the RNA-DDX19 interaction may be a new rate-limiting step when the auto-inhibitory helix is not present. Finally, we observed very modest additional stimulation by Gle1^{CTD}•Nup42^{GBM} and RNA combined over RNA-mediated stimulation alone in the helix mutants, which indicates that Gle1^{CTD}•Nup42^{GBM} does not have a major role outside relief of auto-inhibition. On the other hand, in DDX19^{WT}, the combined effect of Gle1^{CTD}•Nup42^{GBM} and RNA is significantly higher than either alone (Figure 3b). Thus, we conclude that Gle1 activates DDX19 by allowing formation of an open state where the auto-inhibitory helix is displaced.

Nup214 stimulates DDX19 in the presence of Gle1

Nup214 is a cytoplasmic filament protein that binds DDX19 via the same interface as RNA, so it inhibits RNA-mediated stimulation of DDX19, as described previously. In addition, it has been shown that Gle1 can rescue this inhibition in yeast and return the enzyme to fully stimulated levels as achieved by Dbp5^{WT} (von Moeller et al., 2009; Napetschnig et al., 2009; Noble et al., 2011). We observed a similar effect with the human proteins, but we also observed additional stimulation by Nup214 in the combined presence of Gle1^{CTD}•Nup42^{GBM}, RNA, and Nup214 (Figure 3b). We found a modest concentration

dependent stimulatory effect by Nup214 in the presence of Gle1, which was amplified in limiting amounts of $Gle1^{CTD}$•$Nup42^{GBM}$. Furthermore, we observed that Nup214 inhibited the hyperactive DDX19 mutants to levels of activity similar to $DDX19^{WT}$ (Figure 3c). This apparent contradiction of Nup214 being both stimulatory and inhibitory may be reconciled in a model where Nup214 aids Gle1 in separating the RecA domains of DDX19, as crystallographic data shows the interface for Nup214 binding is between the RecA domains of DDX19 (Napetschnig et al., 2009). In this case, Nup214 accelerates formation of the open state, but it must be removed for RNA to bind the enzyme and form the catalytically competent closed state. Further studies are required to determine the rate-limiting steps and precise order of events in the DDX19 catalytic cycle.

Gle1 disease mutants are not deficient in DDX19 activation

We mapped the LAAHD/LCCS1/ALS point mutations to the crystal structure of $hsGle1^{CTD}$•$Nup42^{GBM}$•$DDX19^{\Delta N53}$, but the mutations did not appear to affect $Gle1^{CTD}$ binding to either $Nup42^{GBM}$ or DDX19. In comparing the ability of the mutants to stimulate DDX19 ATPase activity relative to wild-type, there was no significant difference between the mutant $Gle1^{CTD}$•$Nup42^{GBM}$ proteins and wild-type (Figure 4).

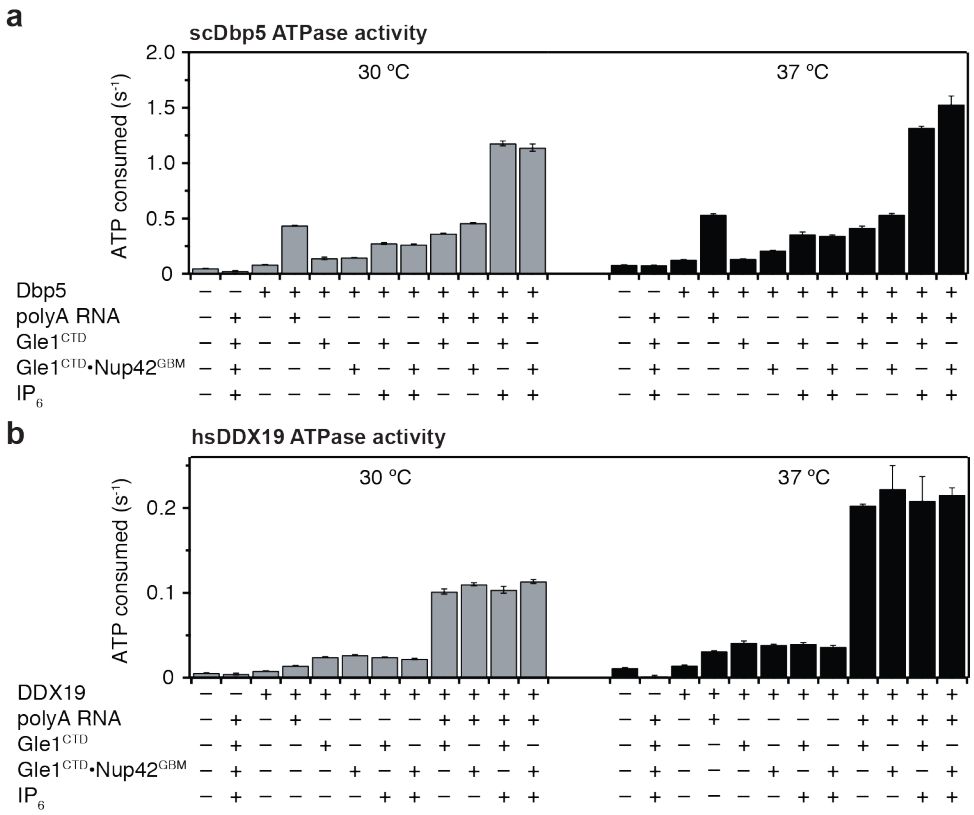


Figure 1. Steady-state ATPase activity of DDX19. (a) Steady state colorimetric ATPase assay with *S. cerevisiae* Dbp5 performed at 30°C and 37°C with either purified *S. cerevisiae* $Gle1^{CTD}$ or $Gle1^{CTD}$•$Nup42^{GBM}$. Reactions were performed with 0.5 μM Dbp5, 1.0 μM $Gle1^{CTD}$ or $Gle1^{CTD}$•$Nup42^{GBM}$, 0.1 mg/ml polyA RNA, and 2.0 μM IP_6. Values shown are the average of three experiments. Error bars indicate the standard deviation. (b) Steady state colorimetric ATPase assay with *H. sapiens* DDX19 performed at 30 °C and 37 °C with either purified *H. sapiens* $Gle1^{CTD}$ or $Gle1^{CTD}$•$Nup42^{GBM}$. Reactions were performed with 2.5 μM DDX19, 5.0 μM $Gle1^{CTD}$ or $Gle1^{CTD}$•$Nup42^{GBM}$, 0.1 mg/ml polyA RNA, and 10.0 μM IP_6. Values shown are the average of three experiments. Error bars indicate the standard deviation.

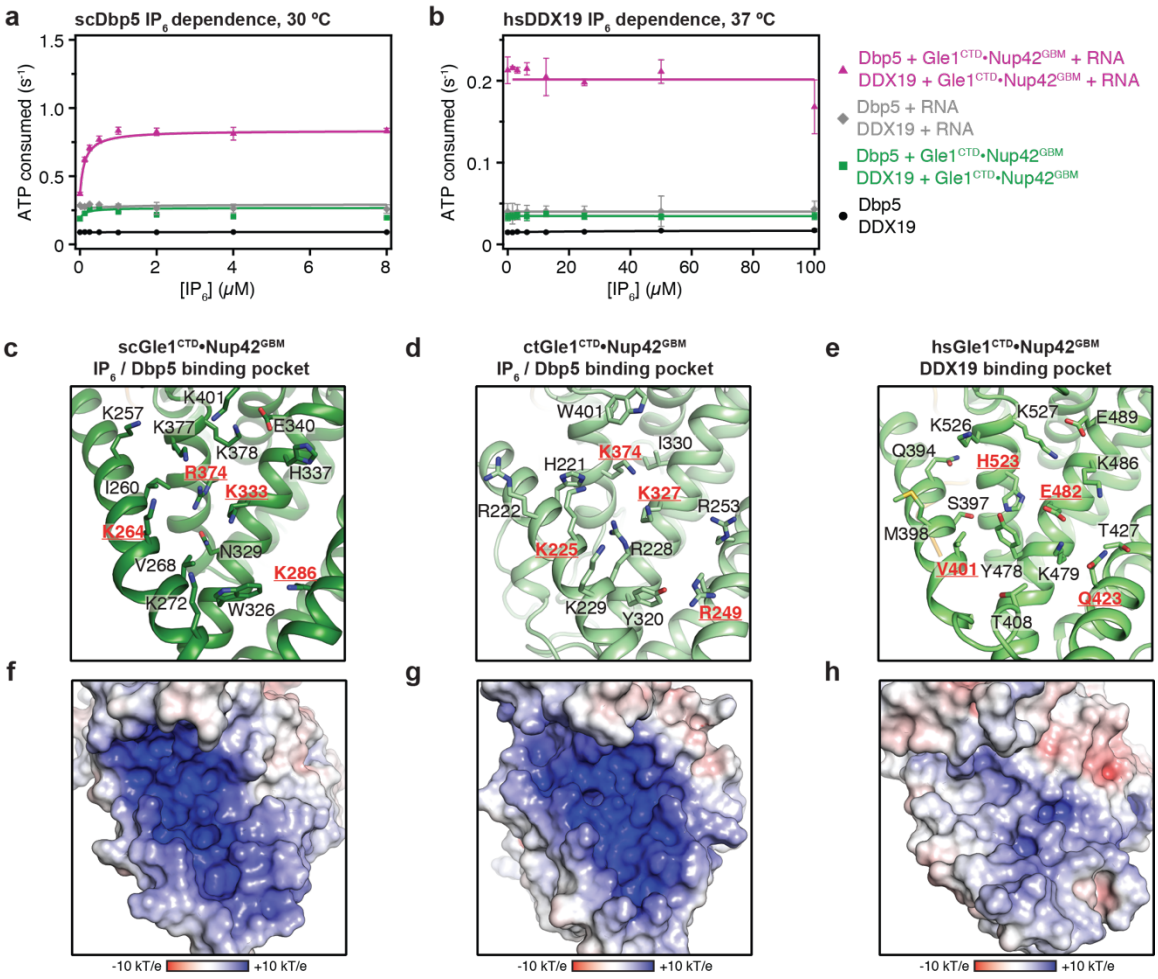

a
scDbp5 IP6 dependence, 30 ºC
ATP consumed (s-1)
[IP6] (μM)
b
hsDDX19 IP6 dependence, 37 ºC
ATP consumed (s-1)
[IP6] (μM)
Dbp5 + Gle1CTD•Nup42GBM + RNA
DDX19 + Gle1CTD•Nup42GBM + RNA
Dbp5 + RNA
DDX19 + RNA
Dbp5 + Gle1CTD•Nup42GBM
DDX19 + Gle1CTD•Nup42GBM
Dbp5
DDX19
c
scGle1CTD•Nup42GBM
IP6 / Dbp5 binding pocket
K257
K401
K377
E340
K378
H337
R374
I260
K333
K264
N329
V268
K286
K272
W326
d
ctGle1CTD•Nup42GBM
IP6 / Dbp5 binding pocket
W401
I330
K374
H221
K327
R222
R253
K225
R228
K229
Y320
R249
e
hsGle1CTD•Nup42GBM
DDX19 binding pocket
K526
K527
E489
Q394
H523
K486
S397
E482
M398
T427
V401
Y478
K479
Q423
T408
f
g
h
-10 kT/e
+10 kT/e

Figure 2: Human Gle1^{CTD} activation of DDX19 is IP_6 independent. Experimental data presented in panels d and g of this figure have been conducted by A.R.C. (a) IP_6 dependence of *S. cerevisiae* performed at 30 °C with purified *S. cerevisiae* Gle1^{CTD}•Nup42^{GBM}. Reactions were performed with 0.5 µM Dbp5, 1.0 µM Gle1^{CTD}•Nup42^{GBM}, 0.1 mg/ml polyA RNA, and the indicated amounts of IP_6. Values shown are the average of three experiments. Error bars indicate the standard deviation. (b) IP_6 dependence of *H. sapiens* DDX19 activation performed at 37 °C with purified *H. sapiens* Gle1^{CTD}•Nup42^{GBM}. Reactions were performed with 2.5 µM DDX19, 5.0 µM Gle1^{CTD}•Nup42^{GBM}, 0.1 mg/ml polyA RNA, and the indicated amounts of IP_6. Values shown are the average of three experiments. Error bars indicate the standard deviation. (c-e) Zoom view of the IP_6 binding pocket of (c) *S. cerevisiae*, (d) *C. thermophilum*, and (e) *H. sapiens* Gle1^{CTD}. Residues that are conserved in fungi but not metazoans are highlighted in red. (f-h) Surface electrostatic potential analysis of IP_6 binding pockets for (f) *S. cerevisiae*, (g) *C. thermophilum*, and (g) *H. sapiens* Gle1. The same view as (c-e) is shown in surface representation colored by electrostatic potential, from red (-10 k_BT/e) to white (0 k_BT/e) to blue (+10 k_BT/e), revealing a dramatically reduced electrostatic potential for human Gle1^{CTD}.

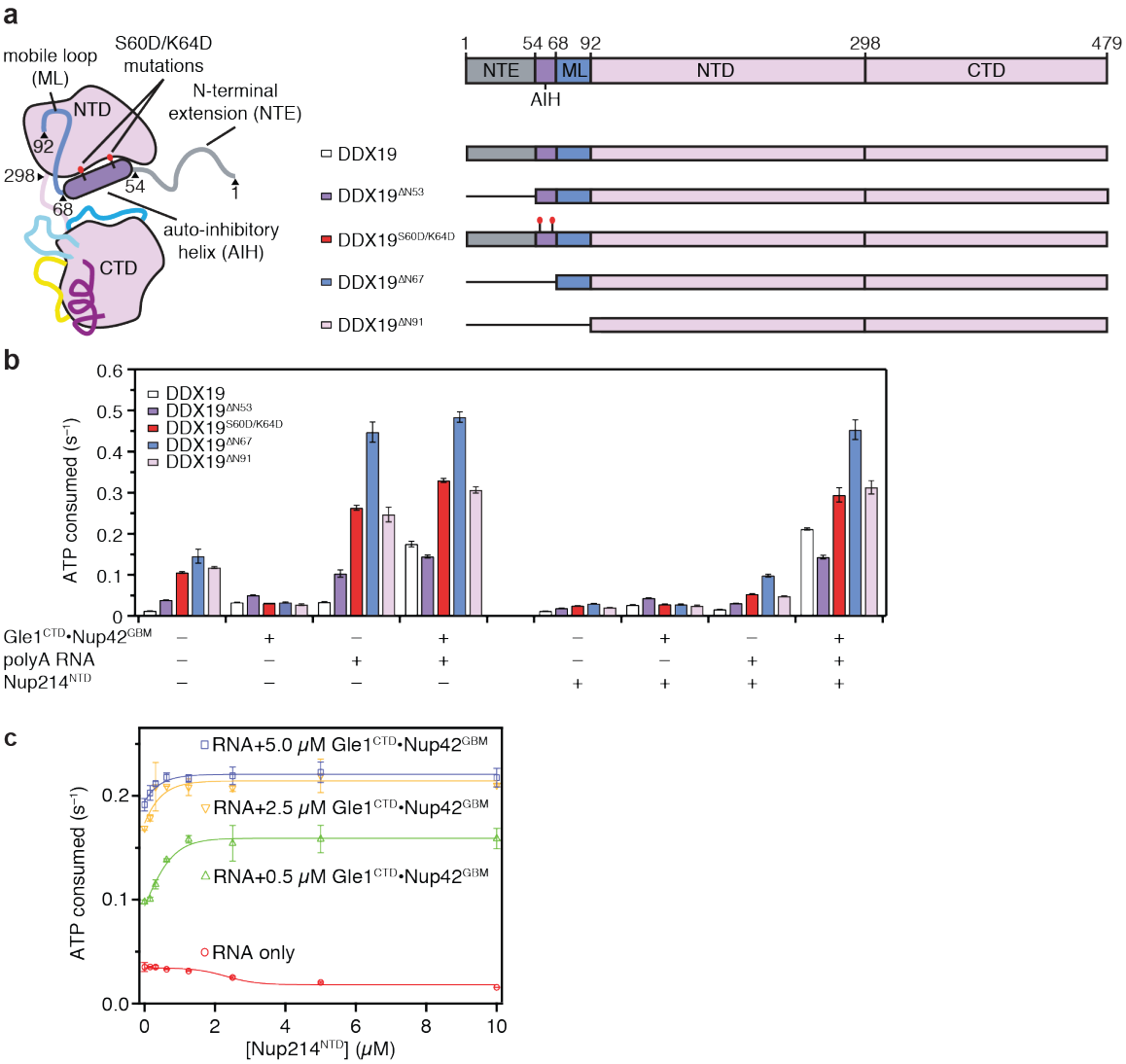


Figure 3. Biochemical analysis of DDX19 auto-inhibition and Nup214-mediated stimulation. (a) Schematics of the construct boundaries of DDX19 variants used in steady-state ATPase assays. (b) Steady-state ATPase activity of DDX19 variants in the presence and absence of RNA, Gle1^{CTD}•Nup42^{GBM}, and Nup214^{NTD}. Values reported are the average of three experiments. Error bars represent standard deviation. (c) Analysis of concentration dependence of Nup214^{NTD} on DDX19 stimulation in the presence of RNA and Gle1^{CTD}•Nup42^{GBM}.

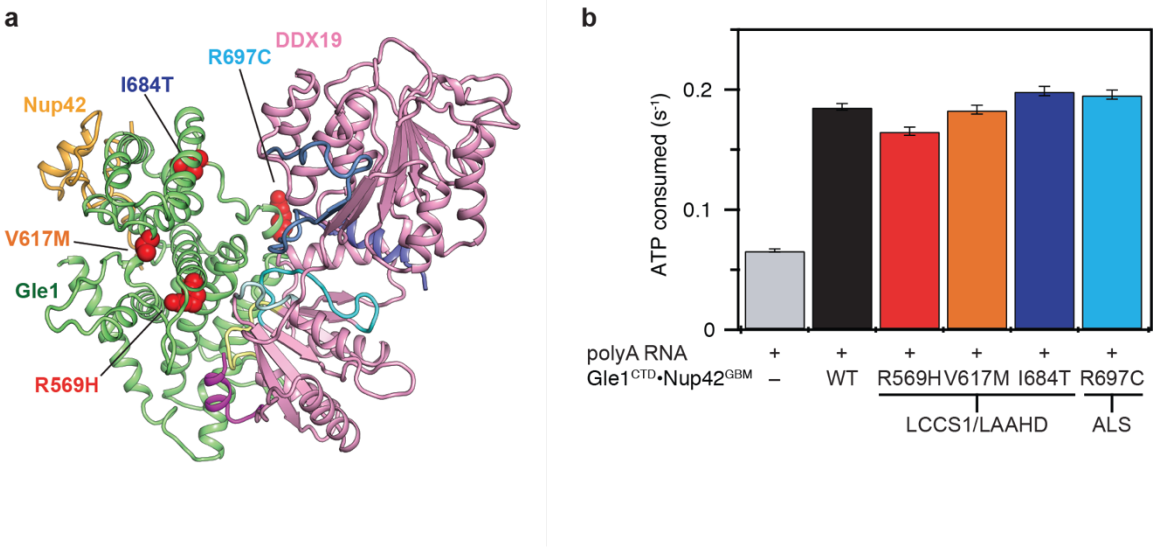


Figure 4. Effects of disease mutations on DDX19 ATPase activity. (a) Mapping single amino acid substitutions associated with human diseases LCCS1/LAAHD and ALS onto the structure of Gle1^{CTD}•Nup42^{GBM}•DDX19$^{\Delta N53}$(ADP). Coloring is the same as in Figure S7a. (b) Steady-state ATPase activity of DDX19 stimulation by Gle1^{CTD} variants containing single amino acid substitutions associated with human disease. All reported values are the average of three experiments. Error bars indicate standard deviation.

C h a p t e r 3

IN VITRO TRANSCRIPTION AND PURIFICATION OF CTE-RNA

Introduction

In eukaryotic cells, mRNA is synthesized in the nucleus from DNA, where the process is spatially separated from translation by ribosomes in the cytoplasm (Hoelz et al., 2011). After post-translational processing such as splicing and addition of a poly(A)+ tail, the mature mRNA is packaged with proteins into an export-competent messenger ribonucleoprotein (mRNP) complex (Stewart, 2010). Export serves as a checkpoint to ensure RNA fidelity; because only fully processed RNAs should be exported, the cell must have mechanisms to prevent the premature export of unprocessed mRNPs (Grüter, 1998). For example, the binding of splicing factors to the mRNP prevents export, sequestering it until splicing is completed (Legrain and Rosbash, 1989). Thus, the removal of specific retention factors that associate with mRNPs facilitates translocation through the NPC. One such export factor that mediates the passage of the mRNP through the NPC is the heterodimeric export factor NXF1•NXT1, which interacts with and shuttles the mRNP through the phenylalanine-glycine (FG) repeats that compose the NPC diffusion barrier (Grüter, 1998; Katahira et al., 1999). At the cytoplasmic face of the NPC, the mRNP encounters the DDX19 export machinery that facilitates specific removal of NXF1•NXT1 and release of the mRNA into the cytoplasm. Removal of NXF1•NXT1 guides unidirectional passage, ensuring the mRNA cannot re-enter through the diffusion barrier (Lund and Guthrie, 2005). Whereas the main steps in how the NPC regulates mRNA export by DDX19 activation are well characterized, as presented earlier in this text, the precise mechanism for removal of NXF1•NXT1 from the mRNP by DDX19 remains to be determined.

NXF1•NXT1 has been studied fairly extensively, along with its yeast homolog Mex67•Mtr2. NXF1•NXT1 is believed to serve as one of the last checkpoints to create an export-competent mRNP, and specific adaptor proteins are required to load the complex on to a processed transcript before it can be translocated. The precise mechanism by which NXF1•NXT1 can recognize the mature transcripts is unclear, although it is thought that the adaptor proteins remodel the mRNP to facilitate NXF1•NXT1 association (Stewart, 2010).

While NXF1•NXT1 is an important factor in bulk mRNA export, it is also utilized by unspliced retroviral RNA. Retroviral replication depends on the export of viral RNAs, which can be incompletely spliced or unspliced, to encode viral proteins and to serve as the genomic RNA for virions assembling from the cytoplasm. Thus, viral RNA must have evolved methods of bypassing the standard cell checkpoints designed to block export of unspliced transcripts. Complex retroviruses, such as HIV, encode the viral Rev protein with a conventionally spliced transcript. The Rev protein can then interact with the Rev recognition element (RRE) sequence of unspliced RNAs encoded by the viral late genes and facilitate nuclear export via the Rev-NES pathway (Fischer, 1995; Hope, 1997). On the other hand, simple retroviruses do not express the Rev protein. Instead, simple retroviruses have a *cis*-acting RNA element, the constitutive transport element (CTE), at the 3' end of their single stranded RNA genome that allows direct binding to NXF1•NXT1 without the adaptor proteins required for bulk mRNA export. Thus, by hijacking the NXF1•NXT1, the viral RNA can bypass the earlier fidelity checks and successfully translocate out of the NPC. This CTE is analogous in function to the RRE, but it uses cellular factors and the bulk mRNA export pathway, instead of Rev and the CRM1 export pathway. The CTE, identified in

Simian type-D retroviruses, is a ~130 nucleotide long element that forms a twofold symmetric RNA stem-loop motif (Hope, 1997; Grüter et al., 1998; Aibara et al., 2015).

In the current literature, there are several crystal structures of most domains of NXF1•NXT1, including a structure of NXF1 bound to one symmetrical half of the CTE-RNA (Aibara et al., 2015; Teplova et al., 2011). An atomic-resolution structure of the CTE-bound heterodimer will provide a framework for understanding how NXF1•NXT1 interacts with RNA and is a starting point for further studies in whether CTE-RNA also utilizes the DDX19 machinery and in how DDX19 specifically remodels RNA structure. Here, we describe a method for large-scale synthesis and purification of CTE-RNA for crystallography.

Development of an *in vitro* transcription protocol for CTE-RNA

To study the interaction between CTE-RNA and NXF1•NXT1, a purification protocol for the CTE is required. Five versions of the CTE were generated: CTE FL is 125 nt long and contains the full extended stem-loop, and the truncation constructs remove bases in a symmetric manner at the opposite end of the hairpin. CTE V1 is 87 nt long, CTE V3 is 77 nt long, CTE ½ is 63 nt long, and CTE ½-2 is 61 nt long (one nucleotide removed at either end of CTE ½). The transcription template also contains *cis*-acting ribozymes to ensure specific cleavage of the 5' and 3' ends of the CTE, with a hammerhead (HH) ribozyme (57 nt) flanking the 5' end, and a Hepatitis Delta Virus (HDV) ribozyme (85 nt) flanking the 3' end. Purified CTE transcripts were desired, so a method of purification to remove the ribozymes was developed. Due to the similarities in size of the ribozymes and some CTE constructs,

additional constructs were developed to contain loop structures in the ribozymes. These loops did not affect catalysis, but they increased the ribozyme sizes to allow for better separation of the CTEs. Constructs are summarized in Table S3

Initial tests of CTE transcription yielded incompletely cleaved transcripts. The hammerhead ribozyme appeared to cleave inefficiently, and the desired CTE transcript was not obtained (Figure 1b, lane 1). In previous work, it has been shown that the HH ribozyme requires at least 5 additional nucleotides complementary to the target, 5' to the first catalytically required sequence (CUGA), to hybridize with the target RNA for cleavage (Price et al., 1995; Ferré-D'Amaré and Doudna, 1996) (Figure 1a). The existing CTE constructs were mutagenized with a 6-nt insertion of the required complementary bases in the HH sequence and transcribed again, this time observing the presence of the fully-cleaved CTE transcript (Figure 1b, lane 2). An additional thermocycling step post-transcription, as described by Lustig et al. (1995), additionally improved ribozyme efficiency, evidenced by the disappearance of the HH-CTE-HDV transcript (Figure 1b, lane 3). Transcripts were obtained for all CTE constructs, with the exception of the CTE ½-2 (Figure 1c).

Scaling of CTE-RNA purification for crystallography

For crystallization trials, milligrams of pure RNA are required. The transcription reaction can be easily scaled up to a greater volume, and thus a greater yield of products, but a large scale purification method remains challenging. The traditional method of RNA preparation for biochemical analyses (which typically require only microgram quantities of RNA) involves loading the transcription reaction mixture on a (40 x 60 x 0.3 cm) preparative polyacrylamide slab gel, and extracting the RNA from gel slices physically cut from the gel

(Heus et al., 1990; Cunningham et al., 1996). This method is both inefficient and limited in resolution of specific RNA products, so a method described by Cunningham et al. (1996) was modified for large-scale purification of the CTE transcripts for crystallography.

Adapting a Model 491 Prep Cell (BioRad), the RNA transcription mixture was loaded on a cylindrical polyacrylamide gel, which was continuously electrophoresed as a peristaltic pump ran elution buffer over the gel, through a UV-vis detector, and into fractions. An ÄKTA Start Protein Purification System (GE Healthcare) was adapted to connect to the Model 491 Prep Cell, serving as the peristaltic pump, detecting the UV-260 trace, and collecting fractions. An initial test run of this method was conducted with a 1 ml transcription reaction containing the CTE FL construct. The elution was performed at a conservative 0.5 ml/min, compared to the 1 ml/min described by Cunningham et al. (1996), on a native 8% TBE polyacrylamide gel. The system ran for 20 hours, yielding separation shown in Figure 1d.

From this initial test, several observations were made. (1) The CTE FL construct was too close in size to the HDV ribozyme, so the resolution of the gel was not high enough to obtain pure CTE FL (Figure 1d, red box). (2) The extended length of time required to run the system resulted in additional ribozyme cleavage of initially uncleaved transcripts occurring as the transcription reaction moved through the gel. Evidence of this additional cleavage is observed the later lanes of Figure 1d where additional HH ribozyme elutes after the larger HDV and CTE transcripts. (3) A transcript the size of the HH ribozyme is separable and relatively pure in the early elution fractions (Figure 1d), which suggests that the CTE ½ construct is a promising target, especially with the looped HH construct isolating the CTE ½

further (magenta boxes in Figure 1c). With this method to increase the scale of RNA purification, enough pure CTE-RNA can be obtained for X-ray crystallography studies.

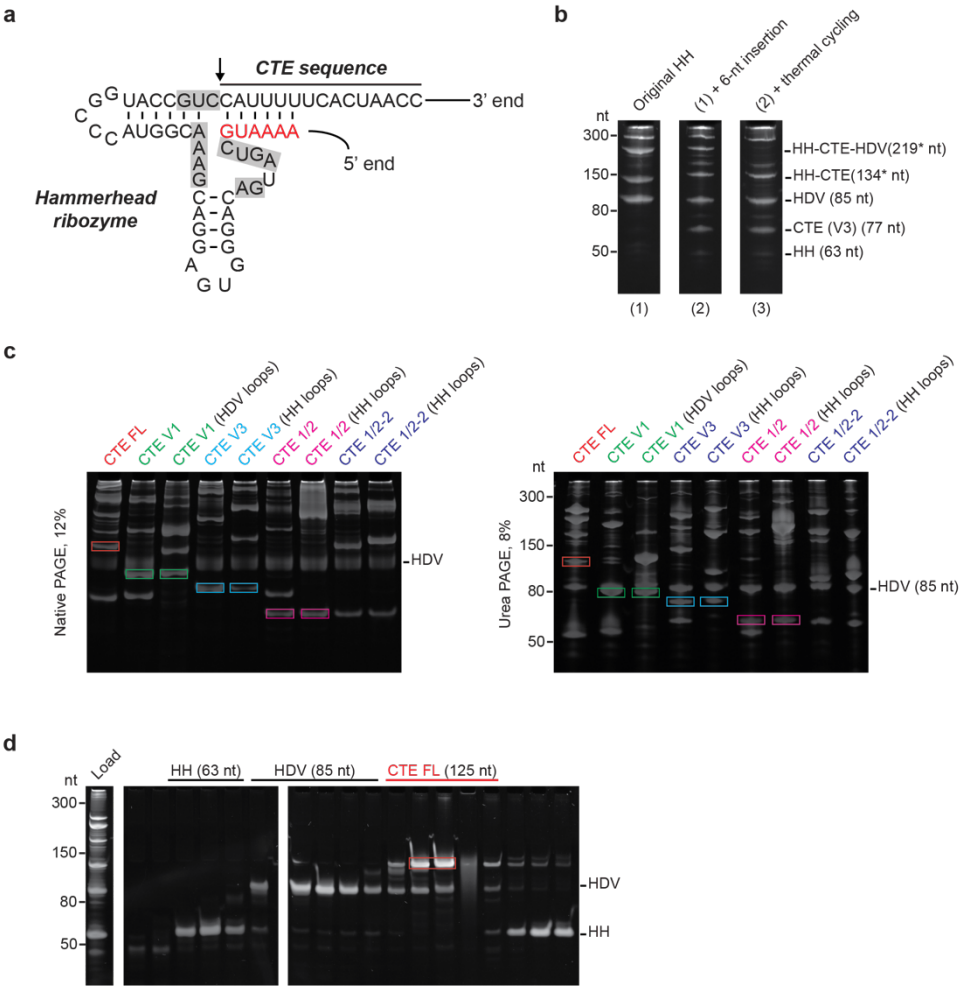

a
CTE sequence
3' end
5' end
Hammerhead ribozyme
b
Original HH
(1) + 6-nt insertion
(2) + thermal cycling
nt
300
150
80
50
HH-CTE-HDV(219* nt)
HH-CTE(134* nt)
HDV (85 nt)
CTE (V3) (77 nt)
HH (63 nt)
(1)
(2)
(3)
c
CTE FL
CTE V1
CTE V1 (HDV loops)
CTE V3
CTE V3 (HH loops)
CTE 1/2
CTE 1/2 (HH loops)
CTE 1/2-2
CTE 1/2-2 (HH loops)
Native PAGE, 12%
HDV
Urea PAGE, 8%
HDV (85 nt)
d
Load
HH (63 nt)
HDV (85 nt)
CTE FL (125 nt)
HDV
HH

Figure 1. Scaling *in vitro* transcription for crystallization trials. (a) Schematic of hammerhead ribozyme cleavage structure. Highlighted regions are sequences required for efficient cleavage. The cleavage site is marked with an arrow, and the red sequence indicates the 6-nt insertion complementary to the target sequence. The 5'-end additionally contains the T7 RNA polymerase promoter sequence, and the 3'-end contains the remainder of the CTE sequence and Hepatitis Delta Virus ribozyme. The CTE sequence shown is CTE V3. (*) denotes length before 6-nt insertion. (b) Comparison of *in vitro* transcription products (1) without the 6-nt insertion, (2) with the 6-nt insertion, and (3) with an additional thermal cycling step. Transcription reactions were visualized with ethidium bromide stained 10% denaturing Urea PAGE gels. (c) Transcription products of all CTE versions, visualized with ethidium bromide stained 12% native PAGE and 8% denaturing Urea PAGE gels. Bands corresponding to the ribozyme-cleaved CTE products are highlighted in colored boxes. (d) Fractions eluted from 8% native PAGE gel cast in a Model 491 BioRad Prep Cell 37 mm chamber. RNA was visualized with SYBR Gold stained 8% denaturing Urea PAGE gels. The desired CTE product is highlighted with a red box.

CONCLUSION

Here we have presented a structural and functional analysis of mRNA export regulation by the NPC. First, we show that the interaction between $Gle1^{CTD}$ and $Nup42^{GBM}$ is highly conserved with atomic resolution X-ray crystal structures of the complex from three separate species. The hydrophobic interaction interface revealed by the structures provides a chemical explanation for how $Nup42^{CTD}$ drastically alters $Gle1^{CTD}$ thermostability and, in cells, stabilizes Gle1 at the NPC during heat shock. Clearly, Nup42 plays a large role in maintaining the proper fold of $Gle1^{CTD}$, which also explains its apparent effect on DDX19 activation. Rather than directly stimulating DDX19, $Nup42^{GBM}$ allosterically activates the enzyme by ensuring the stability of $Gle1^{CTD}$. Interestingly, variants of $Gle1^{CTD}$ with point mutations linked to the motor neuron diseases LAAHD and LCCS1 display altered thermostability. Misfolded Gle1 would likely impair the mRNA export cycle, and our results provide a framework for further studies in understanding the molecular basis of human disease linked to Gle1.

We also present the reconstitution of fully stimulated human DDX19, *in vitro*. The biochemistry described here supplements a comprehensive set of X-ray crystal structures displaying the various activation states of DDX19. From these analyses, we found several striking differences between the metazoan and fungal machinery, where in humans, Gle1-mediated DDX19 activation is IP_6 independent and largely relies on Gle1-mediated removal of an auto-inhibitory helix. From a comparison of *apo* $DDX19^{\Delta N53}$ and $Gle1^{CTD}$•$Nup42^{GBM}$•$DDX19^{\Delta N53}$, we have elucidated the precise molecular mechanism of how $Gle1^{CTD}$ activates

the enzyme. By studying the ATPase activity of DDX19 variants mutated in the N-terminal auto-inhibitory helix, we show that $Gle1^{CTD}$ stimulates DDX19 via relief of auto-inhibition. With these results, we propose a new working model of mRNA export by the nuclear pore complex (Figure 1). In our model, DDX19 remains predominantly in a closed state with the auto-inhibitory helix tightly bound, but Gle1 (stabilized by Nup42) can open the enzyme for nucleotide exchange. Here, Nup214 can further accelerate this process. The ATP bound enzyme can then cycle between open and closed states until it encounters NXF1•NXT1 bound mRNA, where DDX19 then displaces the export factor and hydrolyzes ATP to release mRNA and start a new cycle.

One outstanding question not answered by our model is how DDX19 specifically remodels the NXF1•NXT1-bound mRNP to dissociate the export factors. One hypothesis suggests that RNA is forced into a specific secondary structure by adaptor proteins in the nucleus, and this structure is recognized by NXF1•NXT1. When DDX19 contacts the RNA, it may remodel the RNA such that the NXF1•NXT1-compatible structure is removed, thereby removing the NXF1•NXT1. However, this will remain only a hypothesis until further studies of the specific interactions between NXF1•NXT1, RNA, and DDX19 are conducted. Some clues may come from the study of the retroviral CTE-RNA, which is sufficient to interact with NXF1•NXT1 and hijack the bulk mRNA export pathway. We presented a method to synthesize and purify crystallographic quantities of CTE-RNA. With the construct boundaries from a preliminary X-ray crystal structure of NXF1•NXT1•CTE ½, future work would center on obtaining an atomic-resolution structure of the complex and on investigating the interplay between NXF1•NXT1, CTE/bulk mRNA, and DDX19.

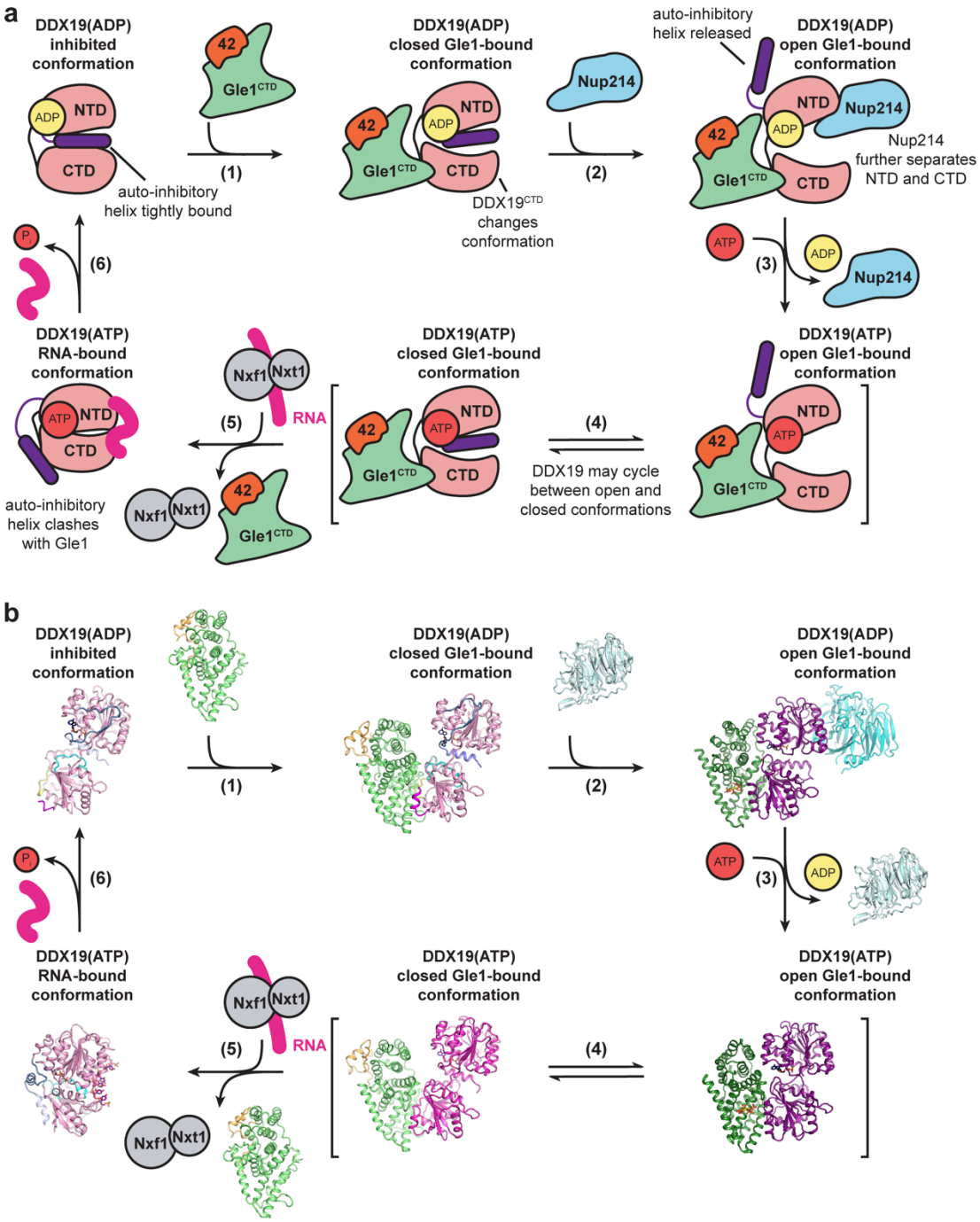

a
DDX19(ADP)
inhibited
conformation
ADP
NTD
CTD
auto-inhibitory
helix tightly bound
42
Gle1CTD
(1)
DDX19(ADP)
closed Gle1-bound
conformation
DDX19CTD
changes
conformation
Nup214
(2)
auto-inhibitory
helix released
DDX19(ADP)
open Gle1-bound
conformation
Nup214
further separates
NTD and CTD
ATP
(3)
Pi
(6)
DDX19(ATP)
RNA-bound
conformation
auto-inhibitory
helix clashes
with Gle1
Nxf1
Nxt1
RNA
(5)
DDX19(ATP)
closed Gle1-bound
conformation
(4)
DDX19 may cycle
between open and
closed conformations
DDX19(ATP)
open Gle1-bound
conformation
b

Figure 1. Proposed working model for DDX19 catalytic cycle. (a) Schematic cartoon of the DDX19 catalytic cycle. (b) Schematic of the DDX19 catalytic cycle with crystal structures of each state. The inhibited conformation corresponds to the crystal structure of $DDX19^{\Delta N53}$(ADP) (PDB ID 3EWS) (Collins et al., 2009). The closed Gle1-bound conformations correspond to the crystal structures of $Gle1^{CTD}$•$Nup42^{GBM}$•$DDX19^{\Delta N53}$(AMP-PNP•Mg^{2+}) and $Gle1^{CTD}$•$Nup42^{GBM}$•$DDX19^{\Delta N53}$(ADP). The Gle1-bound open conformations correspond to the yeast $Gle1^{CTD}$•IP_6•$Dbp5^{\Delta N90}$(ADP) and $Gle1^{CTD}$•IP6•$Dbp5^{\Delta N90}$(ADP)•$Nup159^{NTD}$ (PDB IDs 3RRM and 3RRN) (Montpetit et al., 2011). The RNA bound conformation corresponds to the crystal structure of $DDX19^{\Delta N53}$(AMP-PNP•Mg^{2+})•U_{10} RNA (PDB ID 3G0H) (Collins et al., 2009). $Gle1^{CTD}$•$Nup42^{GBM}$ and $Nup214^{NTD}$ correspond to the *apo* structures (PDB ID 2OIT) (Napetschnig et al., 2009). (Adapted from Lin et al., 2018)

MATERIALS AND METHODS

Construct generation. DNA fragments were amplified using the polymerase chain reaction. SUMO-tagged proteins were cloned into a modified pET28a vector (Mossessova and Lima, 2000). GST-tagged proteins were cloned into a pGEX-6P-1 vector. His_6-tagged proteins were cloned into a modified pET28a vector (Hoelz et al., 2003). S. cerevisiae and H. sapiens $Gle1^{CTD}$•$Nup42^{GBM}$ were cloned into a modified pETDuet-1 vector with $Nup42^{GBM}$ in the first site and $Gle1^{CTD}$ in the second site (Stuwe et al., 2015). Mutants were generated using the QuikChange mutagenesis protocol and confirmed by DNA sequencing. Details of bacterial expression constructs and expression conditions are shown in Supplementary Table 1.

Protein expression and purification. Proteins were expressed in E. coli BL21-CodonPlus(DE3)-RIL cells (Stratagene) in Luria-Bertani media and induced at an OD_{600} of ~0.6 with 0.5 mM IPTG (Gold Biotechnology). Chemicals were purchased from Sigma unless otherwise noted. Unless otherwise noted, cells were harvested by centrifugation and resuspended in a buffer containing 20 mM TRIS (pH 8.0), 500 mM NaCl, 4 mM 2-mercaptoethanol (β-ME), and 15 mM imidazole, supplemented with complete EDTA-free protease inhibitor cocktail (Roche) and flash frozen in liquid nitrogen. Cells were supplemented with 1 mg deoxyribonuclease I (Roche), lysed with a cell disruptor (Avestin), and centrifuged at 4 °C and 30,000 g for 1 hour. Supernatants were loaded onto a Ni-NTA affinity column (GE Healthcare) equilibrated in a buffer containing 20 mM TRIS (pH 8.0),

500 mM NaCl, 4 mM β-ME, and 15 mM imidazole and eluted with a linear gradient of imidazole concentration to 500 mM. Eluted proteins were dialyzed overnight with a buffer containing 20 mM TRIS (pH 8.0), 100 mM NaCl, 4 mM β-ME, and 15 mM imidazole and subsequently purified through affinity, ion exchange, and size exclusion chromatography.

hsGle1^{CTD}•His$_6$-hsNup42^{GBM} variants, scGle1^{CTD}•His$_6$-scNup42^{GBM}, and His$_6$-scGle1^{CTD} were grown for 18 hours at 18 °C. His$_6$-ctGle1^{CTD} was co-expressed with GST-ctNup42^{GBM} and grown for 3 hours at 37 °C. After elution from the Ni-NTA column, the His$_6$ tag and the GST tag were removed by cleavage with PreScission protease concurrent with dialysis. Dialyzed protein was loaded onto a Ni-NTA column equilibrated in a buffer containing 20 mM TRIS (pH 8.0), 100 mM NaCl, 4 mM β-ME, and 15 mM imidazole. Protein-containing fractions were pooled and loaded onto a HiTrap Heparin HP column (GE Healthcare) equilibrated in a buffer containing 20 mM TRIS (pH 8.0), 100 mM NaCl, and 5 mM dithiothreitol (DTT) and eluted with a linear gradient of NaCl concentration to 2 M. Protein-containing fractions were concentrated and loaded onto a HiLoad Superdex 75 16/60 PG column (GE Healthcare) equilibrated in a buffer containing 20 mM TRIS (pH 8.0), 100 mM NaCl, and 5 mM DTT.

His$_6$-SUMO-hsGle1^{CTD} was grown for 18 hours at 18 °C and purified as hsGle1^{CTD}•His$_6$-hsNup42^{GBM}, except the His$_6$-SUMO tag was removed by Ulp1 cleavage.

His$_6$-hsDDX19 variants, His$_6$-scDbp5 and His$_6$-SUMO-ctDbp5 were grown for 18 hours at 18 °C. Cells were harvested by centrifugation and resuspended in a buffer containing 20 mM TRIS (pH 8.0), 500 mM NaCl, 4 mM β-ME, 15 mM imidazole, and 5 % (v/v) glycerol, supplemented with Complete EDTA-free protease inhibitor cocktail (Roche) and flash frozen in liquid nitrogen. After lysis and centrifugation, the supernatant was loaded

onto a Ni-NTA column equilibrated in a buffer containing 20 mM TRIS (pH 8.0), 500 mM NaCl, 4 mM β-ME, 15 mM imidazole, and 5 % (v/v) glycerol. Protein was eluted with a linear gradient of imidazole concentration to 500 mM and protein-containing fractions were dialyzed overnight in a buffer containing 20 mM TRIS (pH 8.0), 100 mM NaCl, 4 mM β-ME, 15 mM imidazole, and 5 % (v/v) glycerol. The His_6 tags of His_6-DDX19 variants and His_6-scDbp5 were cleaved using PreScission protease concurrent with overnight dialysis. The His_6-SUMO tag of His_6-SUMO-ctDbp5 was cleaved with Ulp1 protease and the cleaved protein was desalted with dialysis buffer immediately after elution from the Ni-NTA column. The dialyzed/desalted protein was loaded onto a Ni-NTA column equilibrated in a buffer containing 20 mM TRIS (pH 8.0), 100 mM NaCl, 4 mM β-ME, 15 mM imidazole, and 5 % (v/v) glycerol. Protein-containing flowthrough fractions were loaded onto a HiTrap Q HP column (GE Healthcare) equilibrated in a buffer containing 20 mM TRIS (pH 8.0), 100 mM NaCl, 5 mM DTT, and 5 % (v/v) glycerol and eluted with a linear gradient of NaCl concentration to 2 M. Protein-containing fractions were concentrated and loaded onto a HiLoad Superdex 200 16/60 PG column (GE Healthcare) equilibrated in a buffer containing 20 mM TRIS (pH 8.0), 100 mM NaCl, 5 mM DTT, and 5 % (v/v) glycerol.

His_6-SUMO-hsGle1^N and variants were grown at 30 °C for 2 hours after IPTG induction. After elution from the Ni-NTA column, proteins were dialyzed overnight in a buffer containing 20 mM TRIS (pH 8.0), 100 mM NaCl, and 5 mM DTT. After dialysis, proteins were loaded onto a HiTrap Q HP column and eluted with a linear gradient of NaCl concentration to 2 M. Protein containing fractions were concentrated and loaded onto a Superdex 200 10/300 GL column (GE Healthcare) equilibrated in a buffer containing 20 mM TRIS (pH 8.0), 100 mM NaCl, and 5 mM DTT.

His_6-SUMO-scNup42^{GBM}-His_6 and variants were grown at 37 °C for 2 hours after IPTG induction. After elution from the Ni-NTA column, the His_6-SUMO tag was removed by cleavage with Ulp1 concurrent with overnight dialysis into a buffer containing 20 mM TRIS (pH 8.0), 100 mM NaCl, and 5 mM DTT. After dialysis, proteins were injected onto a HiTrap Q HP column and collected in the flowthrough. Protein-containing fractions were concentrated and loaded onto a Superdex Peptide 10/300 GL column (GE Healthcare).

His_6-Nup214^{NTD} was grown for 18 hours at 18 °C after IPTG induction. After elution from the Ni-NTA column, the protein was dialyzed overnight in 20 mM TRIS (pH 8.0), 500 mM NaCl, 4 mM β-ME, and 15 mM imidazole. The His_6 tag was removed by cleavage with PreScission protease concurrent with dialysis. Dialyzed protein was loaded onto a Ni-NTA column equilibrated in a buffer containing 20 mM TRIS (pH 8.0), 100 mM NaCl, 4 mM β-ME, and 15 mM imidazole. Protein-containing fractions were pooled and loaded onto a HiTrap Q HP column equilibrated in a buffer containing 20 mM TRIS (pH 8.0), 100 mM NaCl, and 5 mM DTT and eluted with a linear gradient of NaCl concentration to 2 M. Protein-containing fractions were concentrated and loaded onto a HiLoad Superdex 200 16/60 PG column equilibrated in a buffer containing 20 mM TRIS (pH 8.0), 100 mM NaCl, and 5 mM DTT.

Crystallization and structure determination. *S. cerevisiae* Gle1^{CTD}•Nup42^{GBM} was crystallized at 23 °C with the hanging drop method using 1 µl of protein solution (22.5 mg/ml) and 1 µl of reservoir solution, containing 0.1 M HEPES (pH 8.2) (Sigma), 11 % (w/v) PEG 3,350, and 0.2 M L-Proline. Crystals were cryoprotected with a solution identical to the reservoir solution, but supplemented with 30 % (v/v) ethylene glycol. X-ray diffraction

data were collected at beamline 12-2 at the Stanford Synchotron Radiation Lightsource (SSRL) and processed with the XDS package (Kabsch, 2010). The structure was solved by molecular replacement with Phaser, using the structure of $Gle1^{CTD}$ (PDB ID 3RRN) as a search model (Montepetit et al., 2011; McCoy et al., 2007). The structure was refined using the PHENIX package with TLS refinement (Adams et al., 2010). One of the Gle1 molecules in the asymmetric unit displays two distinct conformations for half the molecule and was modeled as two alternate conformations for the entire residue range. The final structure was refined to R_{work} and R_{free} values of 18.5 % and 21.1 %, respectively, with excellent geometry (MolProbity score 0.96) (Chen et al., 2010).

H. sapiens $Gle1^{CTD}$•$Nup42^{GBM}$ was crystallized at 23 °C with the hanging drop method using 1 µl of protein solution (20 mg/ml) and 1 µl of reservoir solution, containing 0.2 M sodium potassium phosphate (pH 7.6) and 26 % (w/v) PEG 3,350. Crystals were cryoprotected by gradual supplementation of ethylene glycol in 5 % steps to a final concentration of 30 % (v/v). X-ray diffraction data were collected at beamline 23-ID-D at the Advanced Photon Source (APS) and processed with the XDS package (Kabsch, 2010). The diffraction data was anisotropic, with diffraction limits along the three principal components of 2.8 Å, 3.1 Å, and 3.1 Å. The structure was solved by molecular replacement with Phaser, using the structure of *S. cerevisiae* $Gle1^{CTD}$ (PDB ID 3RRN) as a search model (Montpetit et al., 2011; McCoy et al., 2007). Refinement was performed using torsion NCS restraints with the PHENIX package using anisotropically truncated and scaled data generated using the UCLA-DOE anisotropy server (Adams et al., 2010, Strong et al., 2006). The final structure was refined to R_{work} and R_{free} values of 24.5 % and 27.4 %, respectively,

with excellent geometry (MolProbity score 1.31) (Chen et al., 2010). For details of the data collection and refinement statistics, see Supplementary Table 1.

Yeast strain generation. The *nup42Δ/gle1-GFP* strain was generated in a BY4741 parental strain by first introducing the *natNT2* cassette by homologous recombination into the nup42 gene followed by three rounds of selection on yeast extract peptone dextrose (YPD) plates containing Nourseothricin (Gold Biotechnology) (Janke et al., 2004). Subsequently, a GFP-*kanMX* cassette was inserted into the C-terminus of Gle1 followed by three rounds of selection on YPD plates containing G418 (Janke et al., 2004). Nup42-mCherry variants were introduced using a modified pRS415 plasmid followed by two rounds of selection on plates containing leucine-depleted synthetic dextrose complete medium (SDC-LEU). For details of yeast strains and constructs, see Supplementary Table 3.

Yeast live cell fluorescence. Cells were grown in SDC-LEU medium to mid-log phase at 30 ºC and shifted to 42 ºC for 3 hours. For fluorescence imaging, cells were pelleted by centrifugation for 2 minutes at 650 g, resuspended in water, and imaged using a Carl Zeiss Observer Z.1 equipped with a Hamamatsu camera C10600 Orca-R2.

Yeast growth assay. For growth analysis, cells were grown in SDC-LEU medium at 30 ºC to an OD_{600} of 0.2. 15 µl of a ten-fold dilution series was spotted onto plates containing SDC-LEU medium, which were incubated at 30 ºC and 37 ºC.

GST pull-down interaction analysis: Interaction studies with GST-Nup42 and His_6-Gle1 were performed using GST pull-down experiments. Approximately 100 µl of a 50% (v/v) slurry of glutathione-coupled sepharose beads (GE Healthcare) were equilibrated with a buffer containing 20 mM TRIS (pH 8.0), 100 mM NaCl, and 5 mM DTT. The equilibrated resin was incubated with cleared and filtered lysate from 1 L of bacterial expression culture. The resin was washed 3x with the equilibration buffer (15 ml) and centrifuged with 500 x g at 4°C. Bound protein was eluted from the beads with 250 µl of the wash buffer supplemented with 25 mM reduced glutathione. Eluted protein was resolved by SDS-PAGE and visualized by Coomassie brilliant blue staining.

Differential scanning fluorimetry assay. Differential scanning fluorimetry was performed using a previously described protocol (Niesen et al., 2007). Using a real-time PCR instrument (Bio-Rad C1000 96 well Thermal Cycler), fluorescence of a 20 µl mixture of 5 µM purified protein and 5x SYPRO orange dye (Invitrogen) was measured once per minute while the temperature was increased 1 °C/min from 4 °C to 95 °C. Reactions were performed in 20 mM TRIS (pH 8.0), 100 mM NaCl, and 5 mM DTT. When present, IP_6 was supplemented to 25 µM for yeast proteins and 100 µM for human proteins. The reported T_m values were determined by finding the temperature corresponding to the maximum of the first derivative of the thermal melting curve. Traces represent the average of three experiments.

Pelleting thermostability assay. 50 µl samples of purified protein (10 µg each) were incubated for 30 minutes at indicated temperatures between 25 °C and 55 °C. Soluble and pellet fractions were isolated by centrifugation at 30,000 g for 35 minutes at 4 °C. Reactions

were performed in a buffer containing 20 mM TRIS (pH 8.0), 100 mM NaCl, and 5 mM DTT. When present, IP_6 was added to a final concentration of 20 μM. Proteins were resolved by SDS-PAGE gel and visualized with Coomassie brilliant blue staining.

NADH-coupled ATPase assay. Steady-state ATPase activity rates were determined at 30 °C and 37 °C for scDbp5 and hsDDX19 using previously established conditions (Montpetit et al., 2012). The reaction mixture (80 μl) contained purified scDbp5 and hsDDX19 (WT and variants) at 0.5 μM and 2.5 μM, respectively. Unless otherwise noted, all other S. cerevisiae and H. sapiens proteins ($Gle1^{CTD}$, $Gle1^{CTD}$•$Nup42^{GBM}$, or $Nup214^{NTD}$) were present at concentrations of 1 μM and 5 μM, respectively. When present, polyA RNA with length ranging between 465 and 660 bases (GE Healthcare, 27411001) was added to a final concentration of 0.1 mg/ml. When present, IP_6 was supplemented to a final concentration of 2 μM and 10 μM for yeast and human, respectively, unless otherwise noted. The reaction mixture contained 30 mM HEPES (pH 7.5), 100 mM NaCl, 2 mM $MgCl_2$, 1 mM DTT, 6 mM phosphoenolpyruvate (PEP, Alfa Aesar), 1.2 mM NADH (Sigma), 1.6 μl pyruvate kinase (PK)/lactate dehydrogenase enzyme solution (LDH) (Sigma), and 2.5 mM ATP (Sigma), unless otherwise noted.

For each reaction, two separate mixtures were prepared and kept separate until initiation of the reaction. The first mixture, containing the protein components, IP_6, polyA RNA, and buffer (HEPES, NaCl, and $MgCl_2$), was prepared to a final volume of 20 μl and incubated on ice. The second mixture, containing buffer, DTT, PEP, NADH, ATP, and PK/LDH, was prepared to a final volume of 60 μl and dispensed into a 96-well plate. Reactions were initiated by addition of the protein mixture. Plates were centrifuged at 4,000 g

for 2 minutes at 4 °C prior to being loaded in the pre-warmed plate reader. A_{340} was measured every 30 seconds for 30 minutes using a FlexStation 3 microplate reader (Molecular Devices). Rates were calculated by fitting the linear portion of the reaction. Reported rates were determined by dividing the rate of ATP consumption by the concentration of DDX19/Dbp5 in the reaction. All values reported are the average of three experiments.

RNA Template Generation. CTE constructs were cloned into a modified puc19 vector, flanked by a 5' Hammerhead (HH) ribozyme sequence and a 3' Hepatitis Delta Virus (HDV) ribozyme sequence. Insertions of the 6-nt HH hybridization region were generated using the QuikChange mutagenesis protocol and confirmed by DNA sequencing. Plasmids were digested at 37°C for 16 hours with XbaI and concentrated by ethanol precipitation. Constructs are summarized in Table S3.

In vitro **RNA transcription.** Transcription reactions (100 µl) contained 10 µl 10x T7 transcription buffer, 3 µl 1 M NaOH, 30 µl rNTPs, 5 µg template DNA, 0.4 µl Superase In, 0.25 µl inorganic phosphatase, 5 µl T7 RNA polymerase, and Milli Q water to volume. Reactions were incubated at 37 °C for 4 hours, followed by thermal cycling as previously described (Lustig et al., 1995) and quenching with 2x RNA loading dye. The products were visualized using native TBE-PAGE and denaturing Urea-PAGE and stained with ethidium bromide or SYBR Gold where indicated.

RNA purification. CTE-RNAs were separated on a (40 x 60 x 0.3 cm) preparative Urea-PAGE slab gel, excised after UV-shadowing, and phenol chloroform extracted. For the large-

scale preparation, the transcription reaction mixture was loaded onto a Model 491 Prep Cell (BioRad) 37 mm native TBE-PAGE gel cylinder and electrophoresed. RNA was visualized by UV-detection and fractionated on an ÄKTA-Start. Purified RNAs were visualized using denaturing Urea-PAGE and stained with ethidium bromide or SYBR Gold where indicated.

Illustrations and figures. All structural figures and movies were ray traced in PyMol (Schrödinger). Surface electrostatic potential plots were calculated using Adaptive Poisson-Boltzmann Solver (APBS) (Baker et al., 2001). Secondary structure predictions were generated with the PSIPRED server (Buchan et al., 2013). Sequence alignments were generated using MUSCLE and visualized with ALSCRIPT (Barton, 1993; Edgar, 2004).

REFERENCES

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21.

Adams R.L., Terry, L.J., Wente, S.R. (2014). Nucleoporin FG domains facilitate mRNP remodeling at the cytoplasmic face of the nuclear pore complex. *Genetics* **197**, 1213-1224.

Aibara, S., Katahira, J., Valkov, E., Stewart, M. (2015). The principal mRNA nuclear export factor NXF1:NXT1 forms a symmetric binding platform that facilitates export of retroviral CTE-RNA. *Nucleic Acids Res* **43**, 1883-93.

Alcázar-Román, A.R., Tran, E.J., Guo, S., Wente, S.R. (2006). Inositol hexaphosphate and Gle1 activate the DEAD-box protein Dbp5 for nuclear mRNA export. *Nat Cell Biol* **8**, 711-6.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**, 10037-41.

Barton, G.J. (1993). ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng* **6**, 37-40.

Buchan, D.W., Minneci, F., Nugent, T.C., Bryson, K., Jones, D.T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res* **41**, 349-57.

Chen, V.B., Arendall, W.B. III, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C. (2009). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21.

Collins, R., Karlberg, T., Lehtiö, L., Schütz, P., van den Berg, S., Dahlgren, L.G., Hammarstöm, M., Weigelt, J., Schüler, H. (2009). The DEXD/H-box RNA helicase DDX19 is regulated by an α-helical switch.

Cunningham, L, Kittikamron, K., Lu, Y. (1996). Preparative-scale purification of RNA using an efficient method which combines gel electrophoresis and column chromatography. *Nucleic Acids Res* **24**, 3647-8.

Dossani, Z.Y., Weirich, C.S., Erzberger, J.P., Berger, J.M., Weis, K. (2009). Structure of the C-terminus of the mRNA export factor Dbp5 reveals the interaction surface for the ATPase activator Gle1. *Proc Natl Acad Sci U S A* **106**, 16251-6.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797.

Ferré-D'Amaré, A.R., Doudna, J.A. (1996). Use of *cis*- and *trans*-ribozymes to remove 5' and 3' heterogeneities from milligrams of *in vitro* transcribed RNA. *Nucleic Acids Res* **24**, 977-8.

Fischer, U., Huber, J., Boelens, W.C., Mattajt, L.W., Lührmann, R. (1995). The HIV-1 Rev Activation Domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. *Cell* **82**, 475-83.

Folkmann, A.W., Noble, K.N., Cole, C.N., Wente, S.R. (2011). Dbp5, Gle1-IP6, and Nup159: a working model for mRNP export. *Nucleus* **2**, 540-8.

Grüter, P., Tabernero, C., von Kobbe, C., Schmitt, C., Saavedra, C., Bachi, A., Wilm, M., Felber, B.K., Izaurralde, E. (1998). TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus. *Mol Cell* **1**, 649-59.

Heus, H.A., Uhlenbeck, O.C., Pardi, A. (1990). Sequence-dependent structural variations of hammerhead RNA enzymes. *Nucleic Acids Res* **18**, 1103-8.

Hoelz, A., Nairn, A.C. & Kuriyan, J. (2003). Crystal structure of a tetradecameric assembly of the association domain of Ca2+/calmodulin-dependent kinase II. *Mol Cell* **11**, 1241-51.

Hoelz, A., Debler, E. W., and Blobel, G. (2011). The structure of the nuclear pore complex. *Annu Rev Biochem* **80**, 613-643.

Hope, T.J. (1997). Viral RNA export. *Chem Biol* **4**, 335-44.

Janke, C., Mageira, M.M., Rathfelder, N., Taxis, C., Reber, S., Maekawa, H., Moreno-Borchart, H., Doenges, G., Schwob, E., Schiebel, E., Knop, M. (2004). A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast* **21**, 947-962, (2004).

Kabsch, W. (2010). Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-32.

Kaneb, H.M., Folkmann, A.W., Belzil, V.V., Jao, L.E., Leblond, C.S., Girard, S.L., Daoud, H., Noreau, A., Rochefort, D., Hince, P., *et al.* (2015). Deleterious mutations in the essential mRNA metabolism factor, hGle1, in amyotrophic lateral sclerosis. *Hum Mol Genet* **24**, 1363-73.

Katahira, J., Sträßer, K., Podtelejnikov, A., Mann, M., Jung, J.U., Hurt, E. (1999). The Mex67p-mediated nuclear mRNA export pathway is conserved from yeast to human. *EMBO J* **18**, 2593-2609.

Kendirgi, F., Barry, D.M., Griffis, E.R., Powers, M.A., Wente, S.R. (2003). An essential role for hGle1 nucleocytoplasmic shuttling in mRNA export.

Kendirgi, F., Rexer, D.J., Alcázar-Román, A.R., Onishko, H.M., Wente, S.R. (2005). Interaction between the shuttling mRNA export factor Gle1 and the nucleoporin hCG1: a conserved mechanism in the export of Hsp70 mRNA. *Mol Biol Cell* **16**, 4304-15.

Kosinski, J., Mosalaganti, S., von Appen, A., Teimer, R., DiGuilio, A.L., Wan, W., Bui, K.H., Hagen, W.J., Briggs, J.A., Glavy, J.S., *et al.* (2016). Molecular architecture of the inner ring scaffold of the human nuclear pore complex. *Science* **352**, 363-365.

Legrain, P., Rosbash, M. (1989). Some *cis*- and *trans*-acting mutants for splicing target pre-mRNA to the cytoplasm. *Cell* **57**, 573-83.

Lin, D.H., Stuwe, T.T., Schilbach, S., Rundlet, E.J., Perriches, T., Mobbs, G., Fan, Y., Thierbach, K., Huber, F.M., Collins, L.N., *et al.* (2016). Architecture of the symmetric core of the nuclear pore. *Science* **352**, aaf1015.

Lund, M.K. and Guthrie, C. (2005). The DEAD-box protein Dbp5p is required to dissociate Mex67p from exported mRNPs at the nuclear rim. *Mol Cell* **20**, 645-651.

Lustig, B., Lin, N.H., Smith, S.M., Jernigan, R.L., Jeang, K.T. (1995). A small modified hammerhead ribozyme and its conformational characteristics determined by mutagenesis and lattice calculation. *Nucleic Acids Res* **23**, 3531-8.

McCoy, A.J., Gross-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J. (2007). Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-74.

Montpetit, B., Thomson, N.D., Helmke, K.J., Seeliger, M.A., Berger, J.M., Weis, K. (2011). A conserved mechanism of DEAD-box ATPase activation by nucleoporins and IP6 in mRNA export. *Nature* **472**, 238-242.

Montpetit, B., Seeliger, M.A., Weis, K. (2012). Analysis of DEAD-box proteins in mRNA export. *Methods Enzymol* **511**, 239-254.

Mossessova, E., Lima, C.D. (2000). Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol Cell* **5**, 865-76

Napetschnig, J., Kassube, S.A., Debler, E.W., Wong, R.W., Blobel, G., Hoelz, A. (2009). Structural and functional analysis of the interaction between the nucleoporin Nup214 and the DEAD-box helicase Ddx19. *Proc Natl Acad Sci U S A* **106**, 3089-94.

Niesen, F.H., Berglund, H., Vedadi, M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* **2**, 2212-21.

Noble, K.N., Tran, E.J., Alcázar-Román, A.R., Hodge, C.A., Cole, C.N., Wente, S.R. (2011). The Dbp5 cycle at the nuclear pore complex during mRNA export II: nucleotide cycling and mRNP remodeling by Dbp5 are controlled by Nup159 and Gle1. *Genes Dev* **25**, 1065-77.

Nousiainen, H.O., Kestilä, M., Pakkasjärvi, N., Honkala, H., Kurre, S., Tallila, J., Vuopala K., Ignatius, J., Herva, R., Peltonen, L. (2008). Mutations in mRNA export mediator GLE1 result in a fetal motorneuron disease. *Nat Genet* **40**, 155-7.

Price, S.R., Ito, N., Oubridge, C., Avis, J.M., Nagai, K. (1995). Crystallization of RNA-protein complexes I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. *J Mol Biol* **249**, 398-408.

Rollenhagen, C., Hodge, C.A., Cole, C.N. (2004). The nuclear pore complex and the DEAD box protein Rat8p/Dbp5p have nonessential features which appear to facilitate mRNA export following heat shock. *Mol Cell Biol* **24**, 4869-79.

Schmitt, C., von Kobbe, C., Bachi, A., Panté, N., Rodrigues, J.P., Boscheron, C., Rigaut, G., Wilm, M., Séraphin, B., Carmo-Fonseca, M., Izaurralde, E. (1999). Dbp5, a DEAD-box protein required for mRNA export, is recruited to the cytoplasmic fibrils of the nuclear pore complex via a conserved interaction with CAN/Nup159p. *Embo J* **18**, 4332-47.

Shears, S.B. (2001). Assessing the omnipotence of inositol hexakisphosphate. *Cell Signal* **13**, 151-8.

Stewart, M. (2010). Nuclear export of mRNA. *Trends Biochem Sci* **35**, 609-617.

Strong, M., Sawaya, M.R., Wang, S., Phillips, M., Cascio, D., Eisenberg, D. (2006). Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* **103**, 8060-65.

Stuwe, T.T., Correia, A.R., Lin, D.H., Paduch, M., Lu, V.T., Kossiakoff, A., Hoelz, A. (2015). Architecture of the nuclear pore complex coat. *Science* **347**, 1148-52.

Teplova, M., Wohlbold, L., Khin, N.W., Izaurralde, E., Patel, D.J. (2011). Structure-function studies of nucleocytoplasmic transport of retroviral genomic RNA by mRNA export factor TAP. *Nat Struct Mol Biol* **18**, 990-8.

von Moeller, H., Basquin, C., Conti, E. (2009). DBP5 binds RNA and the cytoplasmic nucleoporin NUP214 in a mutually exclusive manner. *Nat Struct Mol Biol* **16**, 247-54.

Weirich, C.S., Erzberger, J.P., Flick, J.S., Berger, J.M., Thorner, J., Weis, K. (2006). Activation of the DExD/H-box protein Dbp5 by the nuclear-pore protein Gle1 and its coactivator InsP6 is required for mRNA export. *Nat Cell Biol* **8**, 668-76.

APPENDIX

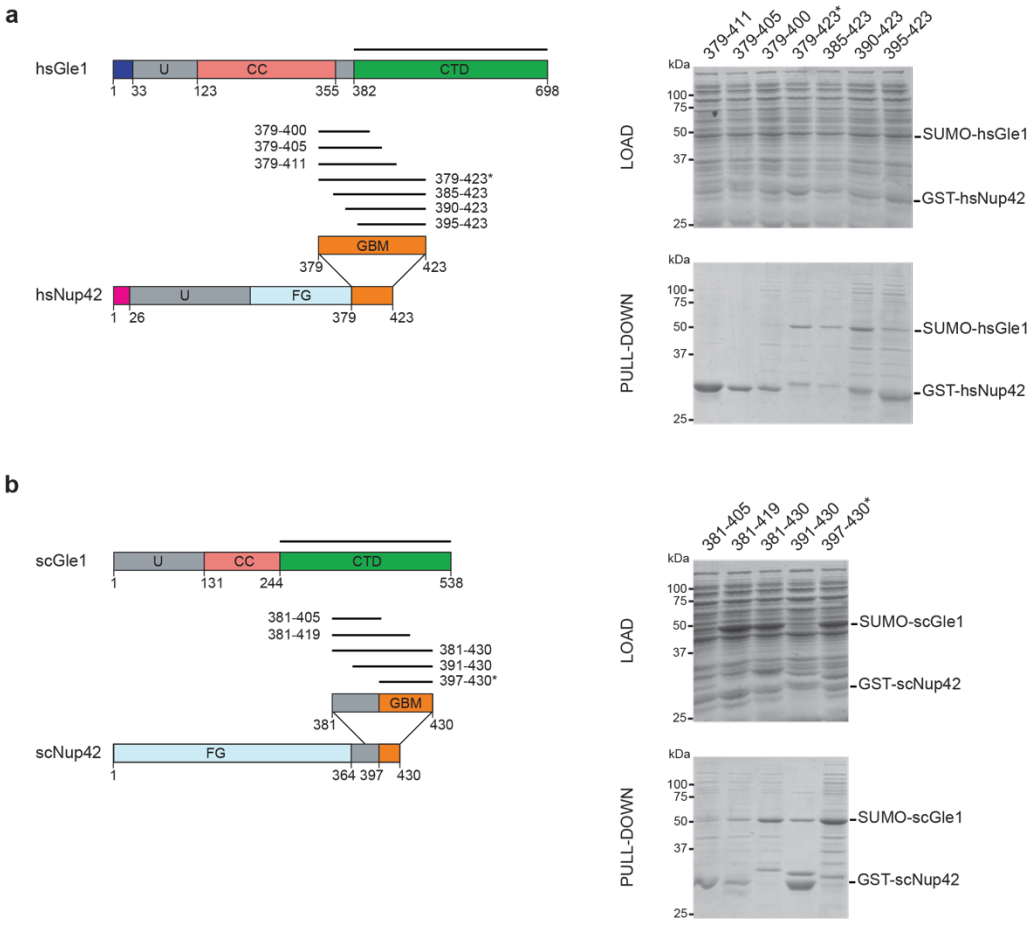


Figure S1. Mapping the minimal Gle1^{CTD} binding region on Nup42. GST-tagged Nup42 fragments and SUMO-tagged Gle1^{CTD} were overexpressed in bacterial culture, and GST-Nup42 was used to pull down SUMO-Gle1^{CTD} on GST resin. (*) indicates crystallized construct. (a) *H. sapiens* Nup42 construct boundaries and pull-down. (b) *S. cerevisiae* Nup42 construct boundaries and pull-down.

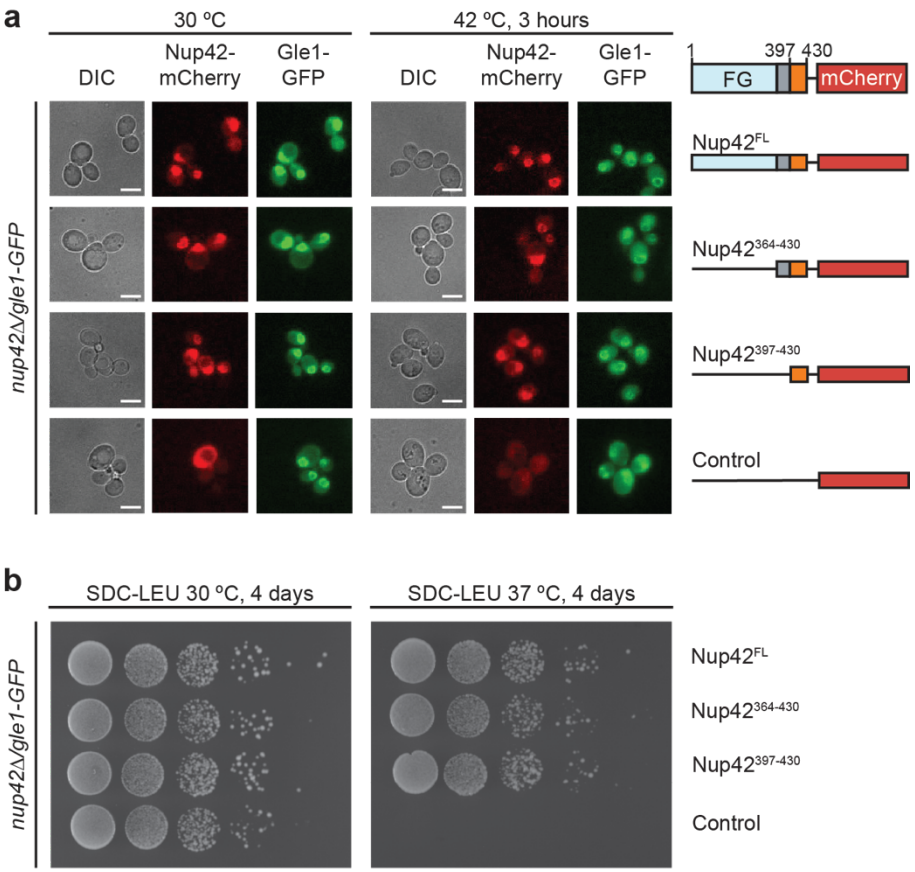
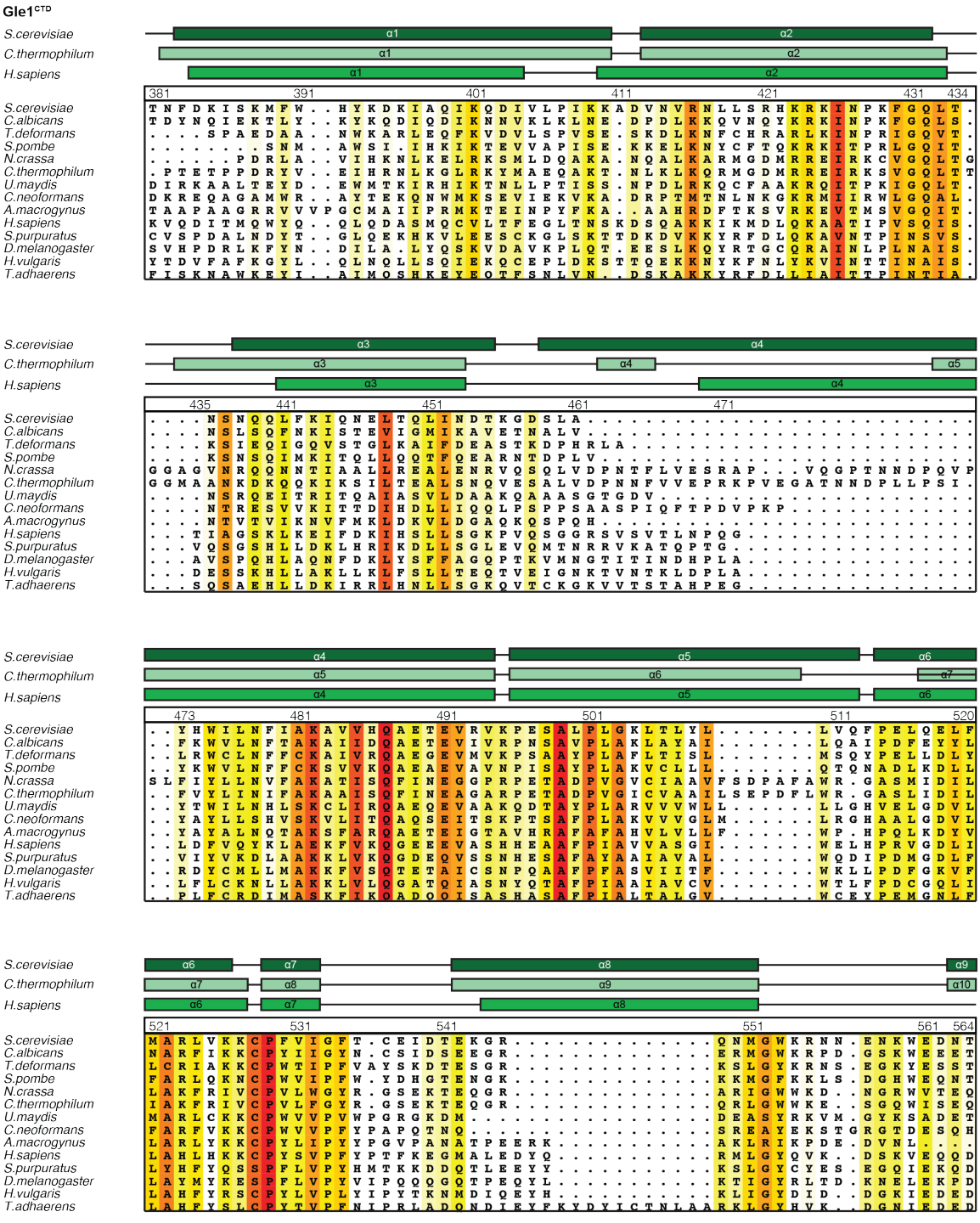


Figure S2. *In vivo* analysis of Nup42^{GBM} and Gle1 heat shock phenotypes. (a) *In vivo* localization analysis in *S. cerevisiae* of Gle1-GFP and Nup42-mCherry variants. Scale bar length is 5 μm. Schematics to the right indicated the fragments of Nup42 that were included in the construct, with omitted fragments indicated by replacement of the domain with a solid line. Residue numbers indicate the fragment included in each construct. (b) Growth analysis of *S. cerevisiae* strains containing the indicated Nup42-mCherry variants. 10-fold serial dilutions were spotted onto SDC-LEU plates and grown for 4 days at 30 and 37°C. Constructs are the same as schematized above.


Gle1CTD
S.cerevisiae
C.thermophilum
H.sapiens
α1
α2
381
391
401
411
421
431
434
S.cerevisiae
C.albicans
T.deformans
S.pombe
N.crassa
C.thermophilum
U.maydis
C.neoformans
A.macrogynus
H.sapiens
S.purpuratus
D.melanogaster
H.vulgaris
T.adhaerens
α3
α4
α5
435
441
451
461
471
α6
α7
473
481
491
501
511
520
α8
α9
α10
521
531
541
551
561
564

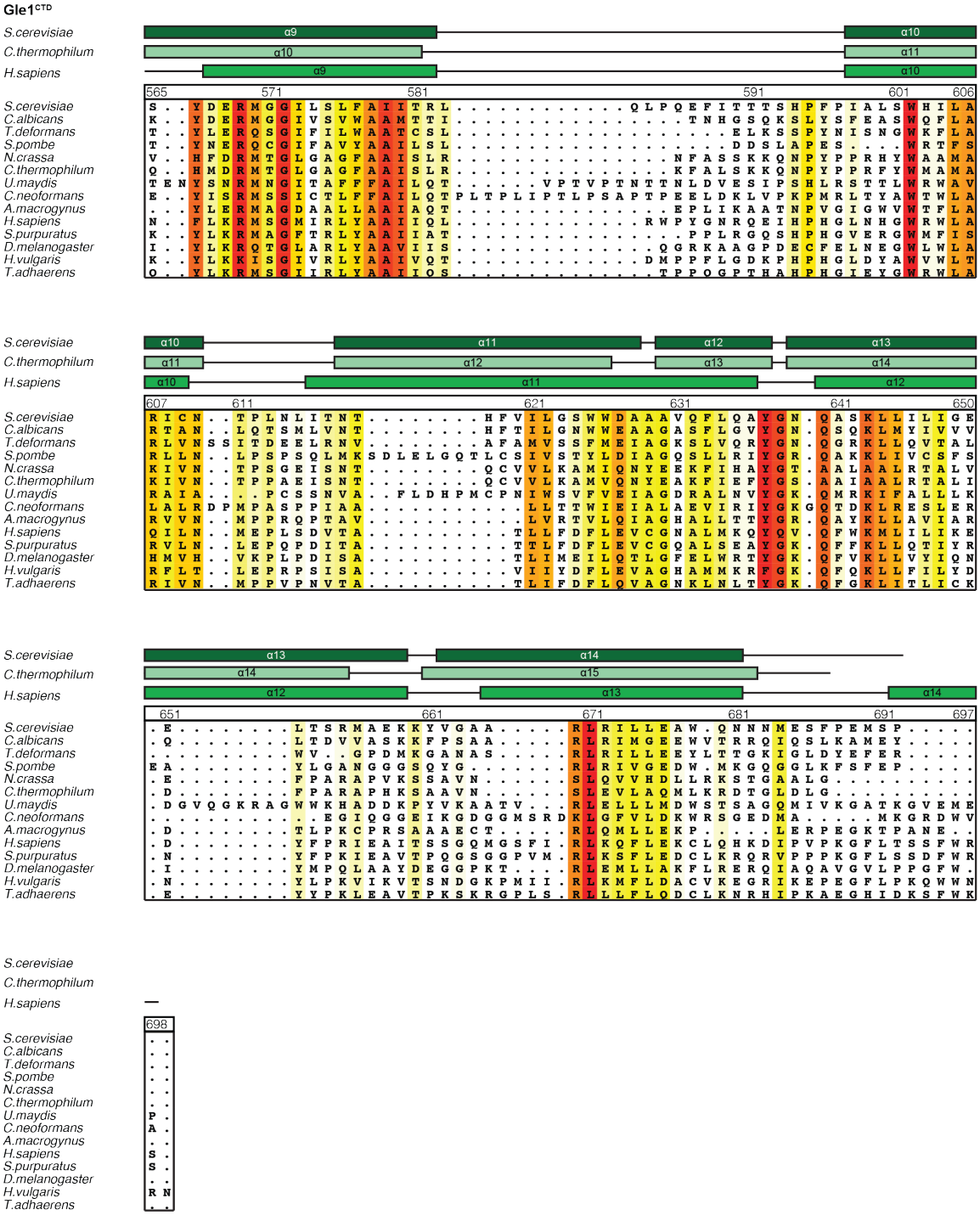

Gle1CTD
S.cerevisiae
C.thermophilum
H.sapiens
α9
α10
α11
α12
α13
α14
α15
565 571 581 591 601 606
607 611 621 631 641 650
651 661 671 681 691 697
698
S.cerevisiae
C.albicans
T.deformans
S.pombe
N.crassa
C.thermophilum
U.maydis
C.neoformans
A.macrogynus
H.sapiens
S.purpuratus
D.melanogaster
H.vulgaris
T.adhaerens

Figure S3. Multispecies sequence alignment of Gle1^{CTD}. This analysis was generated by D.H.L. Sequences from fourteen diverse species were aligned and colored by sequence similarity according to the BLOSUM62 matrix from white (less than 40% similarity), to yellow (55% similarity), to red (100% identity). The numbering is according to the *H. sapiens* protein. The secondary structure is indicated above the sequences as rectangles (α-helices) and lines (unstructured regions) for the *S. cerevisiae, C. thermophilum,* and *H. sapiens* proteins. (Adapted from Lin et al., 2018)

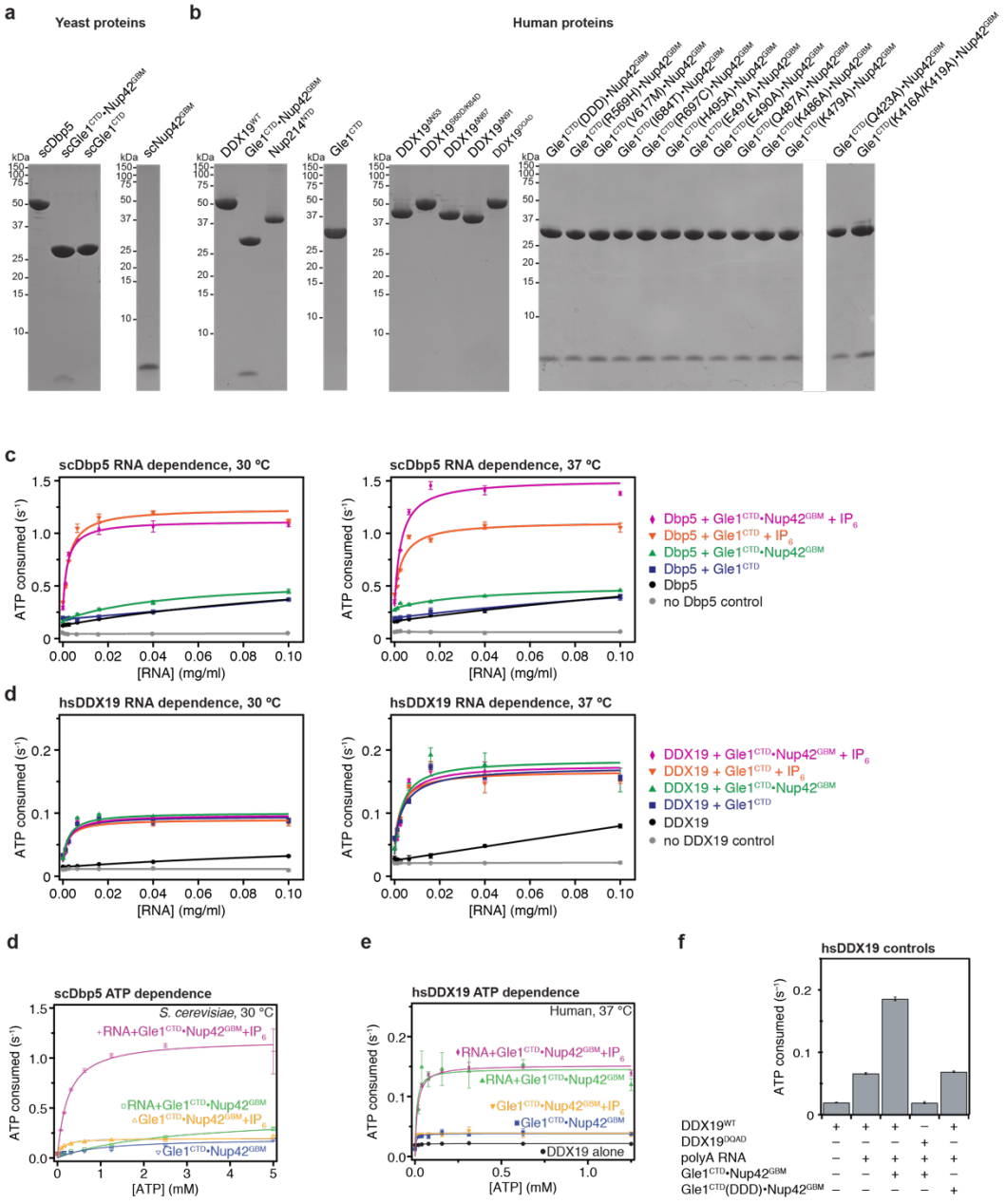

a
Yeast proteins
scDbp5
scGle1CTD•Nup42GBM
scGle1CTD
scNup42GBM
b
Human proteins
DDX19WT
Gle1CTD•Nup42GBM
Nup214FTD
Gle1CTD
DDX19N53
DDX19Q360D/R64D
DDX19N67
DDX19N61
DDX19DQAD
Gle1CTD(DDD)•Nup42GBM
Gle1CTD(R569H)•Nup42GBM
Gle1CTD(V617M)•Nup42GBM
Gle1CTD(I684T)•Nup42GBM
Gle1CTD(R697C)•Nup42GBM
Gle1CTD(H495A)•Nup42GBM
Gle1CTD(E491A)•Nup42GBM
Gle1CTD(E490A)•Nup42GBM
Gle1CTD(Q487A)•Nup42GBM
Gle1CTD(K486A)•Nup42GBM
Gle1CTD(K479A)•Nup42GBM
Gle1CTD(Q423A)•Nup42GBM
Gle1CTD(K416A/K419A)•Nup42GBM
kDa
150
100
75
50
37
25
20
15
10
c
scDbp5 RNA dependence, 30 °C
scDbp5 RNA dependence, 37 °C
ATP consumed (s-1)
[RNA] (mg/ml)
Dbp5 + Gle1CTD•Nup42GBM + IP6
Dbp5 + Gle1CTD + IP6
Dbp5 + Gle1CTD•Nup42GBM
Dbp5 + Gle1CTD
Dbp5
no Dbp5 control
d
hsDDX19 RNA dependence, 30 °C
hsDDX19 RNA dependence, 37 °C
ATP consumed (s-1)
[RNA] (mg/ml)
DDX19 + Gle1CTD•Nup42GBM + IP6
DDX19 + Gle1CTD + IP6
DDX19 + Gle1CTD•Nup42GBM
DDX19 + Gle1CTD
DDX19
no DDX19 control
d
scDbp5 ATP dependence
S. cerevisiae, 30 °C
RNA+Gle1CTD•Nup42GBM+IP6
RNA+Gle1CTD•Nup42GBM
Gle1CTD•Nup42GBM+IP6
Gle1CTD•Nup42GBM
ATP consumed (s-1)
[ATP] (mM)
e
hsDDX19 ATP dependence
Human, 37 °C
RNA+Gle1CTD•Nup42GBM+IP6
RNA+Gle1CTD•Nup42GBM
Gle1CTD•Nup42GBM+IP6
Gle1CTD•Nup42GBM
DDX19 alone
ATP consumed (s-1)
[ATP] (mM)
f
hsDDX19 controls
ATP consumed (s-1)
DDX19WT
DDX19DQAD
polyA RNA
Gle1CTD•Nup42GBM
Gle1CTD(DDD)•Nup42GBM

Figure S4. Steady state colorimetric ATPase activity assay controls. (a, b) SDS-PAGE analysis of the purified (a) yeast or (b) human proteins used in activity assays. (c) RNA dependence of *S. cerevisiae* Dbp5 activation. Reactions were performed at 30 °C and 37 °C with 0.5 μM Dbp5, 1.0 μM $Gle1^{CTD}$ or $Gle1^{CTD}$•$Nup42^{GBM}$, and 2.0 μM IP_6, with the indicated amounts of RNA. Values reported are the average of three experiments. Error bars represent standard deviation. (d) RNA dependence of *H. sapiens* DDX19 activation. Reactions were performed at 30 °C and 37 °C with 2.5 μM DDX19, 5.0 μM $Gle1^{CTD}$ or $Gle1^{CTD}$•$Nup42^{GBM}$, and 10.0 μM IP_6, with the indicated amounts of RNA. Values reported are the average of three experiments. Error bars represent standard deviation. (e) ATP dependence of Dbp5 activity. Values reported are the average of three experiments. Error bars represent standard deviation. (f) ATP dependence of DDX19 activity. Values reported are the average of three experiments. Error bars represent standard deviation. (g) Steady-state ATPase activity of a catalytically dead DDX19 mutant or a Gle1CTD mutant that abolishes stimulation of DDX19 by Gle1CTD, indicating that no activity can be attributed to contaminating factors. Values reported are the average of three experiments. Error bars represent standard deviation. (Adapted from Lin et al., 2018).

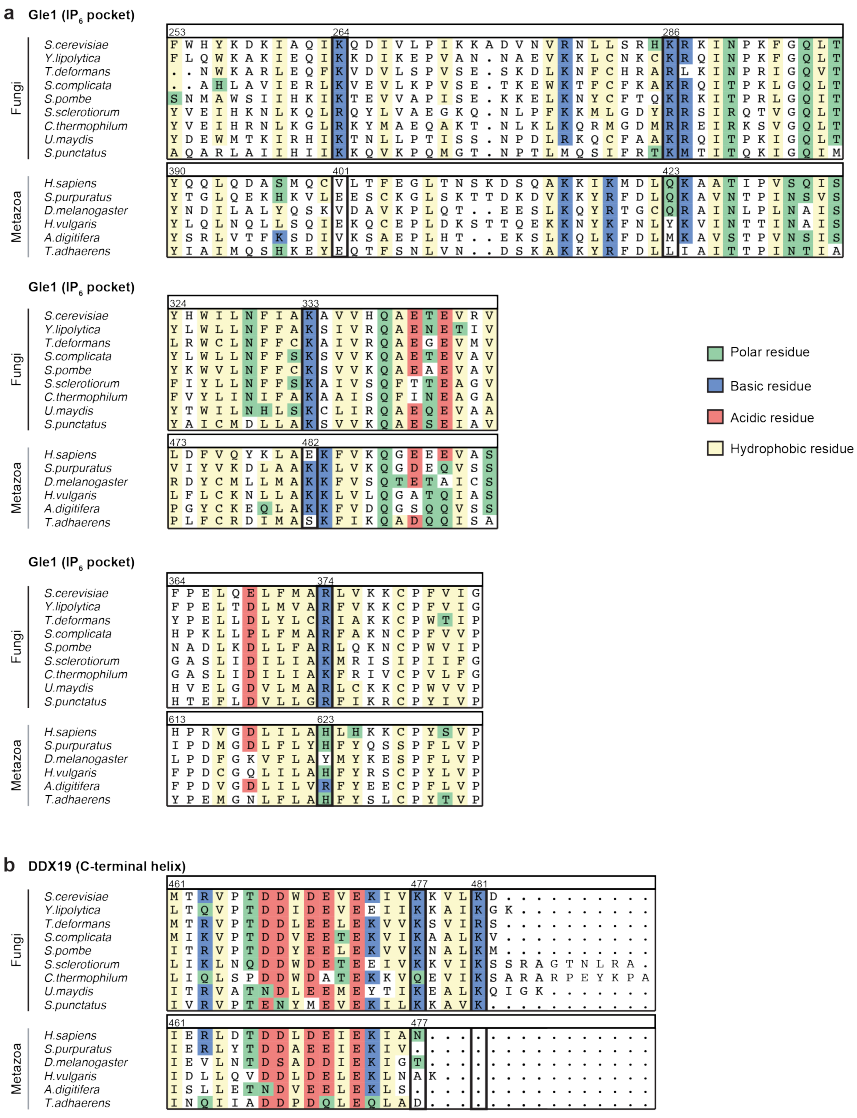


Figure S5. Multispecies sequence alignment of the Gle1 IP_6 pocket and the DDX19 C-terminal helix. This analysis was generated by D.H.L. Basic residues conserved in fungi are outline with black boxes. Residues are colored according to an adapted Clustal color scheme. (Adapted from Lin et al., 2018).

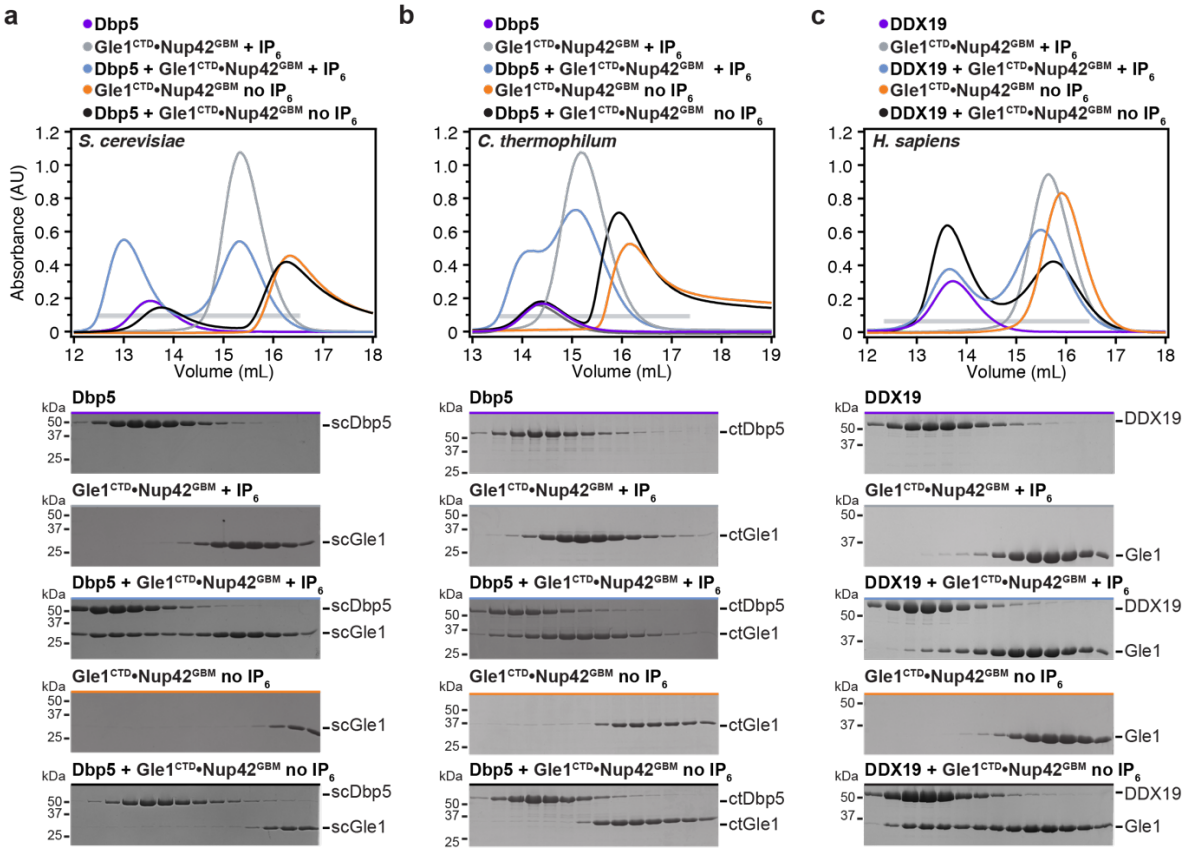


Figure S5. IP_6 dependence of the Gle1^{CTD}•Nup42^{GBM}•DDX19 interaction. This experiment was conducted by D.H.L. (a-c) SEC analysis of the interaction between Dbp5/DDX19 and Gle1^{CTD}•Nup42^{GBM} for (a) *S. cerevisiae*, (b) *C. thermophilum*, and (c) *H. sapiens*, in the presence or absence of IP_6. The elution profiles for Dbp5/DDX19 are shown in purple, Gle1^{CTD}•Nup42^{GBM} are shown in gray or orange, Dbp5/DDX19 with Gle1^{CTD}•Nup42^{GBM} and IP_6 are shown in light blue, and Dbp5/DDX19 with Gle1^{CTD}•Nup42^{GBM} without IP_6 are shown in black. Fungal Gle1^{CTD}•Nup42^{GBM} interacts strongly with the Superdex matrix in the absence of IP_6. The gray horizontal bar indicates fractions visualized with Coomassie-stained SDS-PAGE gels shown below. (Adapted from Lin et al., 2018).

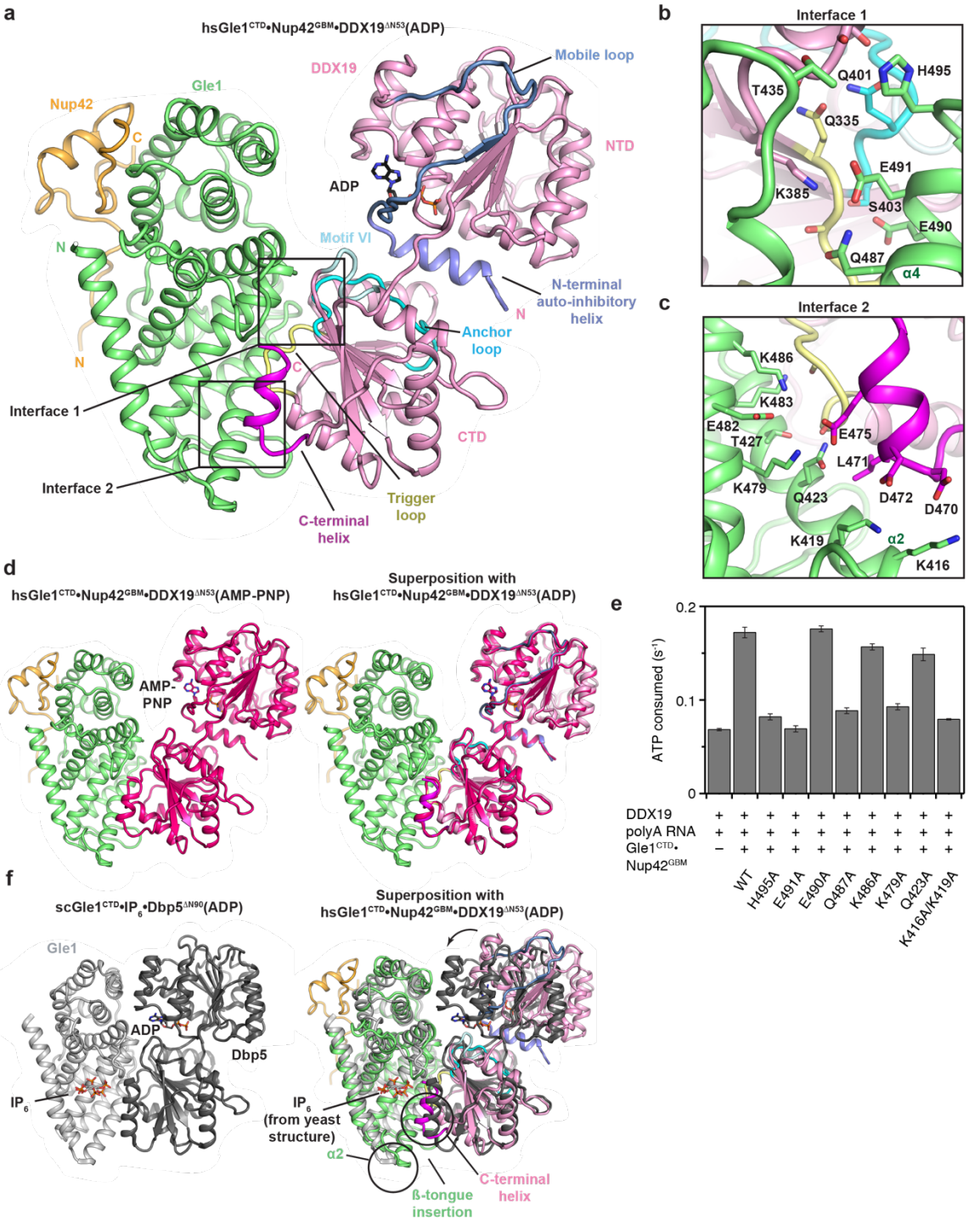

a
hsGle1CTD•Nup42GBM•DDX19ΔN53(ADP)
DDX19
Mobile loop
Gle1
Nup42
C
NTD
ADP
Motif VI
N
N-terminal auto-inhibitory helix
N
Anchor loop
C
Interface 1
CTD
Interface 2
Trigger loop
C-terminal helix
b
Interface 1
Q401
H495
T435
Q335
E491
K385
S403
E490
Q487
α4
c
Interface 2
K486
K483
E482
T427
E475
L471
K479
Q423
D472
D470
K419
α2
K416
d
hsGle1CTD•Nup42GBM•DDX19ΔN53(AMP-PNP)
Superposition with hsGle1CTD•Nup42GBM•DDX19ΔN53(ADP)
AMP-PNP
e
ATP consumed (s-1)
0.2
0.1
0
DDX19
polyA RNA
Gle1CTD•Nup42GBM
WT
H495A
E491A
E490A
Q487A
K486A
K479A
Q423A
K416A/K419A
f
scGle1CTD•IP6•Dbp5ΔN90(ADP)
Superposition with hsGle1CTD•Nup42GBM•DDX19ΔN53(ADP)
Gle1
ADP
Dbp5
IP6
IP6 (from yeast structure)
α2
ß-tongue insertion
C-terminal helix

Figure S7. Structure of the human Gle1^{CTD}•Nup42^{GBM}•DDX19$^{\Delta N53}$ complex. This experiment was conducted by D.H.L. and A.R.C. (a) Crystal structure of *H. sapiens* Gle1^{CTD}•Nup42^{GBM}•DDX19$^{\Delta N53}$(ADP). Motifs of interest are colored and labeled: auto-inhibitory helix (residues 54-67, purple); mobile loop (residues 68-91, blue); trigger loop (residues 328-335, yellow); anchor loop (residues 390-403, cyan); motif VI (residues 429-435, light cyan); C-terminal helix (residues 468-479, magenta). Boxes indicate the regions shown to the right. (b, c) Close-up views of critical complex-forming interactions in interfaces 1 and 2. (d) Left: crystal structure of *H. sapiens* Gle1^{CTD}•Nup42^{GBM}•DDX19$^{\Delta N53}$(AMP-PNP•Mg^{2+}). DDX19 is colored magenta for clarity. Right: Superposition of the ADP and AMP-PNP•Mg^{2+} bound structures. (e) Analysis of the effect of single amino acid substitutions in the Gle1-DDX19 interface on Gle1-mediated stimulation of DDX19, using the same conditions as in Figure 1b at 37 °C. Values shown are the average of three experiments. Error bars indicate standard deviation. (f) Left: crystal structure of *S. cerevisiae* Gle1^{CTD}•IP_6•Dbp5$^{\Delta N90}$(ADP) (PDB ID 3RRN) (Montpetit et al., 2011). Right: superposition of the *S. cerevisiae* and *H. sapiens* structures. The arrow indicates the rotation relating the conformations observed in the two crystal structures. Circles highlight structural differences. (Adapted from Lin et al., 2018).

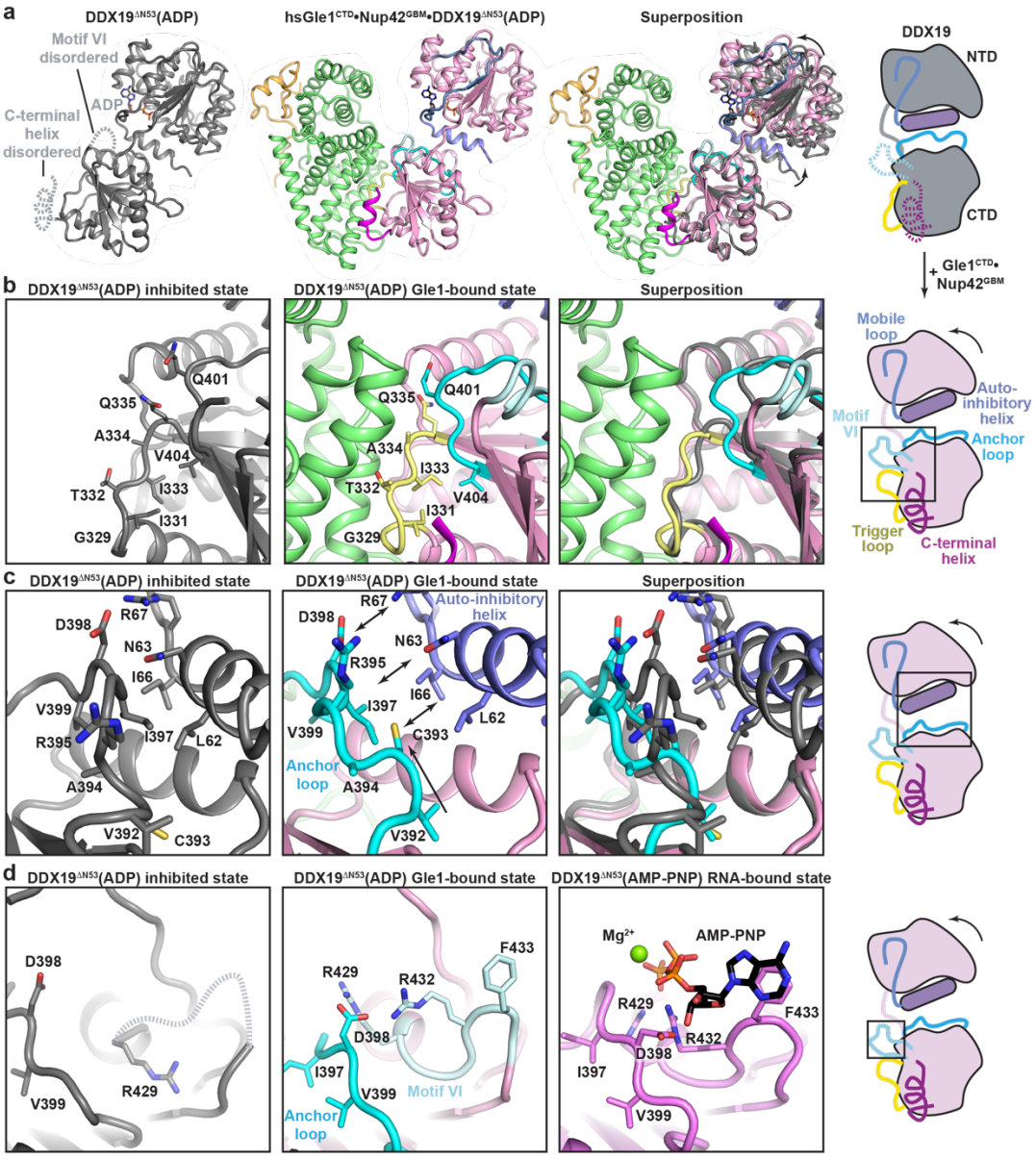

a
DDX19ΔN53(ADP)
hsGle1CTD•Nup42GBM•DDX19ΔN53(ADP)
Superposition
Motif VI disordered
ADP
C-terminal helix disordered
DDX19
NTD
CTD
+ Gle1CTD• Nup42GBM
b
DDX19ΔN53(ADP) inhibited state
DDX19ΔN53(ADP) Gle1-bound state
Superposition
Q401
Q335
A334
V404
T332
I333
I331
G329
Mobile loop
Auto-inhibitory helix
Motif VI
Anchor loop
Trigger loop
C-terminal helix
c
DDX19ΔN53(ADP) inhibited state
DDX19ΔN53(ADP) Gle1-bound state
Superposition
D398
R67
N63
I66
V399
R395
I397
L62
A394
V392
C393
Auto-inhibitory helix
Anchor loop
d
DDX19ΔN53(ADP) inhibited state
DDX19ΔN53(ADP) Gle1-bound state
DDX19ΔN53(AMP-PNP) RNA-bound state
D398
R429
V399
F433
R432
I397
Motif VI
Anchor loop
Mg2+
AMP-PNP

Figure S8: Conformational changes in DDX19 induced by Gle1 binding. This experiment was conducted by D.H.L. and A.R.C. (a) Left: crystal structure of *H. sapiens* DDX19$^{\Delta N53}$(ADP) (PDB ID 3EWS) (Collins et al., 2009). Disordered regions (C-terminal helix and motif VI) are indicated with dashed lines. Middle: crystal structure of *H. sapiens* Gle1^{CTD}•Nup42^{GBM}•DDX19$^{\Delta N53}$(ADP) shown in the same orientation and colored as in Figure S7a. Right: superposition of the two structures. Arrows indicate the rotation relating the conformations of DDX19$^{\Delta N53}$(ADP) in the presence and absence of Gle1^{CTD}•Nup42^{GBM}. The cartoon on the right schematizes the transition from the inhibited state to the Gle1-bound state. (b) Zoom view of the DDX19 trigger loop (yellow) in (left) the inhibited state (PDB ID 3EWS), (middle) the Gle1-bound state, and (right) their superposition. The cartoon on the right indicates the region of DDX19 shown. (c) Zoom view of the DDX19 anchor loop (cyan) and auto-inhibitory helix (purple) in the (left) inhibited state (PDB ID 3EWS), (middle) the Gle1-bound state, and (right) their superposition. The cartoon on the right indicates the region of DDX19 shown. (d) Zoom view of DDX19 Motif VI in (left) the inhibited state (PDB ID 3EWS), (middle) the Gle1-bound state, and (right) the RNA-bound state (PDB ID 3G0H) (Collins et al., 2009). The cartoon on the right indicates the region of DDX19 shown. (Adapted from Lin et al., 2018).

Supplementary Table 1.

Bacterial expression constructs and expression conditions

#	Protein	Residues	Expression vector	Restriction sites 5', 3'	N-terminal overhang	C-terminal overhang	Expression conditions
1	hsGle1 CTD	382-698	pET28a-SUMO	BamHI, NotI	Smt3p-S	-	18 °C / 18 hours
2	hsNup42 CTD (379-400)	379-400	pGex6P-1 PreS	BamHI, NotI	GST-LEVLFQG PHM	-	18 °C / 18 hours
3	hsNup42 CTD (379-405)	379-405	pGex6P-1 PreS	BamHI, NotI	GST-LEVLFQG PHM	-	18 °C / 18 hours
4	hsNup42 CTD (379-411)	379-411	pGex6P-1 PreS	BamHI, NotI	GST-LEVLFQG PHM	-	18 °C / 18 hours
5	hsNup42 CTD (379-423)	379-423	pGex6P-1 PreS	BamHI, NotI	GST-LEVLFQG PHM	-	18 °C / 18 hours
6	hsNup42 CTD (384-423)	384-423	pGex6P-1 PreS	BamHI, NotI	GST-LEVLFQG PHM	-	18 °C / 18 hours
7	hsNup42 CTD (390-423)	390-423	pGex6P-1 PreS	BamHI, NotI	GST-LEVLFQG PHM	-	18 °C / 18 hours
8	hsNup42 CTD (395-423)	395-423	pGex6P-1 PreS	BamHI, NotI	GST-LEVLFQG PHM	-	18 °C / 18 hours
9	scGle1 CTD	244-538	pET28a-SUMO	BamHI, NotI	Smt3p-S	-	18 °C / 18 hours
10	scNup42 CTD (381-405)	381-405	pGex6P-1 PreS	BamHI, NotI	GST-LEVLFQG PHM	-	18 °C / 18 hours
11	scNup42 CTD (381-419)	381-419	pGex6P-1 PreS	BamHI, NotI	GST-LEVLFQG PHM	-	18 °C / 18 hours
12	scNup42 CTD (381-430)	381-430	pGex6P-1 PreS	BamHI, NotI	GST-LEVLFQG PHM	-	18 °C / 18 hours
13	scNup42 CTD (391-430)	391-430	pGex6P-1 PreS	BamHI, NotI	GST-LEVLFQG PHM	-	18 °C / 18 hours
14	scNup42 CTD (397-430)	397-430	pGex6P-1 PreS	BamHI, NotI	GST-LEVLFQG PHM	-	18 °C / 18 hours
15	scGle1 CTD	244-538	pET28a-PreS	NdeI, XhoI	GPHM	-	18 °C / 18 hours
16♦	scGle1 CTD scNup42 GBM	244-538 397-430	pETDuet PreS	NdeI, XhoI BamHI, NotI	M GPSGS	- -	18 °C / 18 hours
17	hsGle1 CTD	382-698	pET28a-SUMO	BamHI, NotI	S	-	18 °C / 18 hours
18♦	hsGle1 CTD hsNup42 GBM	382-698 379-423	pETDuet PreS	NdeI, XhoI BamHI, NotI	M GPSGS	- -	18 °C / 18 hours
19♦	ctGle1 CTD	216-519	pET28a-PreS	NdeI, NotI	GPHM	-	37C °C / 3 hours co-expressed with ctNup42
20♦	ctNup42 GBM	494-558	pGex6P-1 PreS	EcoRI, XhoI	GPHMGS PEF	-	37C °C / 3 hours co-expressed with ctGle1
21	scNup42 GBM	397-430	pET28a-SUMO	BamHI, XhoI	S	YALEHHH HHH	37 °C / 2 hours
22	scNup42 GBM F409A	397-430	pET28a-SUMO	BamHI, XhoI	S	YALEHHH HHH	37 °C / 2 hours

#	Protein	Residues	Expression vector	Restriction sites 5', 3'	N-terminal overhang	C-terminal overhang	Expression conditions
23	scNup42 GBM F414A	397-430	pET28a-SUMO	BamHI, XhoI	S	YALEHHHHHH	37 °C / 2 hours
24	scNup42 GBM L416A	397-430	pET28a-SUMO	BamHI, XhoI	S	YALEHHHHHH	37 °C / 2 hours
25	scNup42 GBM L416R	397-430	pET28a-SUMO	BamHI, XhoI	S	YALEHHHHHH	37 °C / 2 hours
26	scNup42 GBM P423A	397-430	pET28a-SUMO	BamHI, XhoI	S	YALEHHHHHH	37 °C / 2 hours
27	scNup42 GBM F409D	397-430	pET28a-SUMO	BamHI, XhoI	S	YALEHHHHHH	37 °C / 2 hours
28	scNup42 GBM F414D	397-430	pET28a-SUMO	BamHI, XhoI	S	YALEHHHHHH	37 °C / 2 hours
29	scNup42 GBM F409D/F414D	397-430	pET28a-SUMO	BamHI, XhoI	S	YALEHHHHHH	37 °C / 2 hours
30	scDbp5	1-482	pET28a-PreS	NdeI, NotI	GPHM	-	18 °C / 18 hours
31♦	hsDDX19	1-479	pET28a-PreS	NdeI, NotI	GPH	-	18 °C / 18 hours
32	ctDbp5	1-477	pETMCN-SUMO	BamHI, NotI	S	-	18 °C / 18 hours
33♦	hsDDX19 ΔN53	54-479	pET28a-PreS	NdeI, NotI	GPHM	-	18 °C / 18 hours
34	hsGle1 CTD H495A hsNup42 GBM	382-698 379-423	pETDuet PreS	NdeI, XhoI BamHI, NotI	M GPSGS	- -	18 °C / 18 hours
35	hsGle1 CTD E491A hsNup42 GBM	382-698 379-423	pETDuet PreS	NdeI, XhoI BamHI, NotI	M GPSGS	- -	18 °C / 18 hours
36	hsGle1 CTD E490A hsNup42 GBM	382-698 379-423	pETDuet PreS	NdeI, XhoI BamHI, NotI	M GPSGS	- -	18 °C / 18 hours
37	hsGle1 CTD Q487A hsNup42 GBM	382-698 379-423	pETDuet PreS	NdeI, XhoI BamHI, NotI	M GPSGS	- -	18 °C / 18 hours
38	hsGle1 CTD K486A hsNup42 GBM	382-698 379-423	pETDuet PreS	NdeI, XhoI BamHI, NotI	M GPSGS	- -	18 °C / 18 hours
39	hsGle1 CTD K479A hsNup42 GBM	382-698 379-423	pETDuet PreS	NdeI, XhoI BamHI, NotI	M GPSGS	- -	18 °C / 18 hours
40	hsGle1 CTD Q423A hsNup42 GBM	382-698 379-423	pETDuet PreS	NdeI, XhoI BamHI, NotI	M GPSGS	- -	18 °C / 18 hours
41	hsGle1 CTD K416A/K419A hsNup42 GBM	382-698 379-423	pETDuet PreS	NdeI, XhoI BamHI, NotI	M GPSGS	- -	18 °C / 18 hours
42	hsDDX19 ΔN67	68-479	pET28a-PreS	NdeI, NotI	GPHM	-	18 °C / 18 hours
43	hsDDX19 ΔN91	92-479	pET28a-PreS	NdeI, NotI	GPHM	-	18 °C / 18 hours
44	hsDDX19 S60D/K64D	1-479	pET28a-PreS	NdeI, NotI	GPH	-	18 °C / 18 hours
45	Nup214 NTD	1-405	pET28a-PreS	NdeI, NotI	GPH	-	18 °C / 18 hours
46	hsGle1 CTD G666D/I669D/Q673D hsNup42 GBM	382-698 379-423	pETDuet PreS	NdeI, XhoI BamHI, NotI	M GPSGS	- -	18 °C / 18 hours
47	hsGle1 CTD R569H hsNup42 GBM	382-698 379-423	pETDuet PreS	NdeI, XhoI BamHI, NotI	M GPSGS	- -	18 °C / 18 hours
48	hsGle1 CTD V617M hsNup42 GBM	382-698 379-423	pETDuet PreS	NdeI, XhoI BamHI, NotI	M GPSGS	- -	18 °C / 18 hours
49	hsGle1 CTD I684T hsNup42 GBM	382-698 379-423	pETDuet PreS	NdeI, XhoI BamHI, NotI	M GPSGS	- -	18 °C / 18 hours
50	hsGle1 CTD R697C hsNup42 GBM	382-698 379-423	pETDuet PreS	NdeI, XhoI BamHI, NotI	M GPSGS	- -	18 °C / 18 hours
51	hsDDX19 E243Q (DQAD)	1-479	pET28a-PreS	NdeI, NotI	GPH	-	18 °C / 18 hours

♦ Constructs that were used for crystallization

Supplementary Table 2.

Yeast constructs

Plasmid	Protein	Residues	Vector	Restriction Sites 5', 3'	Selection
pRS415-P_{Nop1}-mCherry	N/A	N/A	pRS415	N/A	LEU2
pRS415-P_{Nop1}-NUP42-mCherry	Nup42	1-430	pRS415	NdeI, SpeI	LEU2
pRS415-P_{Nop1}-nup42 (364-430)-mCherry	Nup42	364-430	pRS415	NdeI, SpeI	LEU2
pRS415-P_{Nop1}-nup42 (397-430)-mCherry	Nup42	397-430	pRS415	NdeI, SpeI	LEU2

Supplementary Table 3.

RNA template constructs

Plasmid	nt	Vector	Ribozymes	Modifications
CTE FL	125	puc19	5' HH, 3' HDV	-
CTE V1	87	puc19	5' HH, 3' HDV	-
CTE V1 loops	87	puc19	5' HH, 3' HDV*	2 loops inserted in HDV
CTE V3	77	puc19	5' HH, 3' HDV	-
CTE V3 loops	77	puc19	5' HH*, 3' HDV	2 loops inserted in HH-
CTE ½	63	puc19	5' HH, 3' HDV	-
CTE ½ loops	63	puc19	5' HH*, 3' HDV	2 loops inserted in HH
CTE ½-2	61	puc19	5' HH, 3' HDV	-
CTE ½-2	61	puc19	5' HH*, 3' HDV	2 loops inserted in HH-

*loops inserted